REPORT RESUMES

ED 015 873 24 SE 003 907

EFFECTS OF EXPERIMENTAL PROGRAMS IN MATHEMATICS ON RELEVANT ATTITUDES AND INTERESTS OF NINTH GRADE PUPILS AS MEASURED BY QUESTIONNAIRE INDICES. INTERIM REPORT.

BY- RYAN, JAMES J.
MINNESOTA STATE DEPT. OF EDUCATION, ST. PAUL

REPORT NUMBER BR-5-1028-IR PUB DATE JUL 67
CONTRACT OEC-5-10-051
EDRS PRICE MF-$0.50 HC-$3.04 74P.

DESCRIPTORS- *CURRICULUM DEVELOPMENT, *EVALUATION, *EDUCATIONAL PROGRAMS, *INSTRUCTION, *STUDENT CHARACTERISTICS, *SECONDARY SCHOOL MATHEMATICS, ATTITUDES, ALGEBRA, GRADE 9, INTERESTS, LEARNING DIFFICULTIES, MATHEMATICS, TEXTBOOKS, AIKEN MATHEMATICS INTEREST SCALE, DUTTON MATHEMATICS ATTITUDE SCALE, THE UNIVERSITY OF ILLINOIS COMMITTEE ON SCHOOL MATHEMATICS, BALL STATE, SCHOOL MATHEMATICS STUDY GROUP,

THIS INTERIM REPORT COVERS ONE PHASE OF A BROADER PROGRAM AIMED AT STUDYING THE EFFECTS OF THREE EXPERIMENTAL MATHEMATICS PROGRAMS IN NINTH-GRADE ALGEBRA--BALL STATE, THE UNIVERSITY OF ILLINOIS COMMITTEE ON SCHOOL MATHEMATICS (UICSM), AND SCHOOL MATHEMATICS STUDY GROUP (SMSG)--ON THE ATTITUDES AND INTERESTS PUPILS DEVELOP TOWARD MATHEMATICS. THE SAMPLE CONSISTED OF 37 PAIRS OF NINTH-GRADE ALGEBRA CLASSES, EACH PAIR BEING TAUGHT BY THE SAME TEACHER, AND UTILIZING ONE OF THE THREE EXPERIMENTAL PROGRAMS (E), IN ONE CLASS, CONVENTIONAL METHODS (C) IN THE OTHER. TESTS WERE ADMINISTERED AT THE BEGINNING AND AT THE END OF THE SCHOOL YEAR. TWO INSTRUMENTS, PREVIOUSLY DEVELOPED BY OTHERS, WERE USED TO MEASURE INTEREST IN MATHEMATICS. A QUESTIONNAIRE WAS ALSO DEVELOPED TO MEASURE THESE SPECIFIC ATTITUDE AND INTEREST FACTORS--(1) INTRINSIC INTEREST, (2) PERCEIVED KNOWLEDGE, (3) PERCEIVED UTILITY, (4) EXPERIENCED EASE OR DIFFICULTY OF LEARNING, AND (5) ACHIEVEMENT MOTIVATION. OTHER EVALUATION MEASURES THAT WERE USED WERE (1) SEQUENTIAL TESTS OF EDUCATIONAL PROGRESS (STEP) TESTS, (2) EIGHTH-GRADE MATHEMATICS GRADES, AND (3) CONCURRENT YEAR, NINTH-GRADE MATHEMATICS GRADES. RESULTS, IN GENERAL, SHOWED LITTLE DIFFERENTIAL EFFECT IN PUPIL ATTITUDES AND INTERESTS BETWEEN THE INSTRUCTIONAL PROGRAMS. HOWEVER, A CONSISTENT TENDENCY WAS REPORTED FOR PUPILS IN THE BALL STATE PROGRAM TO DEVELOP LESS POSITIVE ATTITUDES, AND THOSE IN THE UICSM PROGRAM TO DEVELOP MORE POSITIVE ATTITUDES THAN STUDENTS IN THE CONVENTIONAL PROGRAM. ALSO, IT WAS INDICATED THAT STUDENTS EXPERIENCED MORE LEARNING DIFFICULTY IN THE E PROGRAMS THAN DID STUDENTS IN THE C PROGRAMS. A LIST OF 12 REFERENCES IS INCLUDED. (DH)

PA 24

INTERIM REPORT
BR ~~Project No~~. 5-1028
Contract No. OE-5-10-051

EFFECTS OF EX... IMENTAL PROGRAMS
IN MATHEMATICS ON RELEVANT ATTITUDES
AND INTERESTS OF NINTH GRADE PUPILS AS
MEASURED BY QUESTIONNAIRE INDICES.

July, 1967

U.S. DEPARTMENT OF
HEALTH, EDUCATION, AND WELFARE

Office of Education
Bureau of Research

EFFECTS OF EXPERIMENTAL PROGRAMS IN MATHEMATICS ON RELEVANT ATTITUDES AND INTERESTS OF NINTH GRADE PUPILS AS MEASURED BY QUESTIONNAIRE INDICES.

INTERIM REPORT
Project No. 5-1028
Contract No. OE-5-10-051

James J. Ryan

July, 1967

The research reported herein was performed pursuant to a contract with the Office of Education, U.S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

Minnesota National Laboratory
Minnesota State Department of Education
St. Paul, Minnesota

TABLE OF CONTENTS

TABLE OF CONTENTS-cont.

Effects of experimental programs in mathematics on the relevant attitudes and interests of ninth grade pupils as measured by questionnaire indices[1]

James J. Ryan

I. Introduction

The development of positive attitudes and interests with regard to the content of any subject matter area is usually one of the main instructional objectives emphasized by educators and curriculum developers. This is because the pupil's attitudes and interests directly affect, if not represent, the motivational basis for acquisition and performance in connection with the subject matter. Consequently, such outcomes need to be considered when examining the merits of any instructional program.

With respect to newer programs in mathematics, it is reasonable to consider and their developers and proponents have suggested, that in addition to providing a more effective and functional knowledge of mathematics, these programs might contribute more to the development of positive attitudes and interests in mathematics than have the traditional programs.

More positive attitudes and interests toward mathematics might be expected in part because of several characteristics of the newer programs such as: (1) an attempt to provide more powerful general concepts and principles which have broader applicability and require less emphasis on specific manipulative and computational skills of a somewhat monotonous and repetative nature; (2) emphasis on active pupil participation in the learning process and "discovery learning"; (3) presenting problems and concepts in more meaningful and relevant situations and contexts than has been typical of conventional materials.

This study was carried out as part of a project investigating the effects of several recently developed experimental programs in secondary mathematics on the attitudes and interests pupils develop toward mathematics.[2] This facet of the project was focused on providing an assessment of a broad range of possible attitudinal effects which on logical grounds appeared likely to be influenced by the alternate programs and/or related conditions of instruction. These effects were examined using indices measuring both the general affective reactions of pupils as well as more specific attitude and interest components and factors.

[1]The research reported herein was supported in part through a contract (OE-5-10-051, Project 5-1028) with the Office of Education, U.S. Department of Health, Education and Welfare.

Very capable assistance with the data gathering, processing and analysis for this study was provided by Clifford Carlson, Rodney Rosse, David Klemmack, Chaur C. Chen and JoAnn Youngren which is gratefully acknowledged.

[2]Effects of modern and conventional mathematics curricula on pupil attitudes, interests, and perception of proficiency, Office of Education Project 5-1028, James J. Ryan, Project Director.

The procedure followed was to administer questionnaires including the various indices at the beginning and end of the school year to pairs of ninth grade algebra classes taught by the same teachers, one class with one of the several experimental programs, the other with the teacher's conventional program.

Comparisons were made between pupils in classes receiving the alternate instructional programs in terms of their questionnaire responses. Since these effects might vary with other conditions of instruction or pupil or teacher characteristics; pupil sex, measures of achievement and grades in mathematics, pupil judgments of relevant instructional conditions, and teacher experience with, and evaluation of, the experimental programs were considered in the analysis. Also the analysis considered pupil's initial (beginning of year) level of interest and attitudes to determine the effects in terms of change over the period of instruction.

A. Conceptual Considerations

1. Instructional program effects

The main purpose of this study was to determine whether the experimental mathematics programs individually or collectively contributed to the development of differential attitudes toward mathematics in contrast to the conventional programs of instruction. Since the concern was with effects or pupil change over the year in which the different instructional materials were used, the pupil's entering or initial attitudes and attitude relevant characteristics were considered in the analysis design.

Differences observed on an attitude measure between E and C class pupils when the initial characteristics of pupils in the alternate instructional groups are equated, either statistically or otherwise, could be attributed to differences in the instructional conditions or experiences of the separate groups of pupils. Even though the instructional materials represented the main difference between the alternate groups and the design required that the same teachers instruct both experimental and conventional classes, other attitude relevant instructional factors or conditions may have differed in some consistent or systematic way to contribute to or determine attitude differences between experimental and conventional class pupils. Among the factors or conditions that might vary with the instructional materials were those that could have a direct and those that could have an indirect effect upon pupil attitudes. The indirect effects would be those that resulted from changes in other instructional factors or conditions that were more directly affected by the instructional materials being used. Teachers' attitudes toward the materials, their grading practices, or their demands upon pupil performance if they varied between E and C classes could be possible sources of such indirect effects. The direct effects would be those resulting from the pupils' interaction with the materials, as such, relatively independent of other factors in the instructional situation. The assertion of those suggesting that the newer curricula might make a greater contribution to attitudes toward mathematics appears to imply that such outcomes are primarily the result of direct rather than indirect effects. Consequently, in the analysis consideration was given to and a distinction made between factors that might represent direct and indirect effects of the instructional materials. Among the former, data concerning pupil judgements or reactions to the materials were gathered and among the latter, information concerning teacher's attitudes toward the materials and pupil grades were obtained. From a methodological point of view, however, because of their interaction over time and because these effects are not independent, it may be difficult to determine the order of effect or causal sequence for such concomitant factors with respect to attitudinal outcomes.

In addition to the question of the general effects or differences for the experimental programs being examined, there is also the question of possible variations in effects or outcomes under different instructional conditions or for different subgroups of pupils. This is a question of factors or conditions that might moderate the effects of the various instructional materials. Among the factors that were considered as moderator variables were such pupil characteristics as sex and level of mathematics ability as well as the amount of teacher experience with the experimental program. This set of factors should be distinguished from those considered previously as conditions affecting or mediating the attitudinal outcomes in that differences for the latter variables arise during the instructional period while those characterized as moderator variables represent conditions existing prior to instruction. The causal sequence of effect for the moderator variables is more evident while the order or direction of effect for conditions occuring or at least observed during the instructional period can only be inferred in conjunction with certain assumptions.

2. Attitude and interest measures

With respect to the several attitude and interest measures being used, there are some general methodological questions which should be considered in connection with the objectives of the study. The main objective of the study, to determine the effects of the experimental programs on pupil attitudes, presumes the existence of at least a general affective reaction factor which is measurably independent of other logically distinguishable pupil characteristics which have been frequently observed to be major determinants of pupil behavior. General academic ability or proficiency as indicated by achievement tests and grades and academic achievement motivation in the sense of a positive attitude toward school and school achievement are two such individual difference factors which have frequently been found to account for a large amount of variation in pupil behavior in the school situation. Since the variables with which this study was concerned were to be conceptually distinguished from both general and specific levels of ability or proficiency and from attitudes toward school _in general_, it is necessary that the respective measures of these variables exhibit a certain degree of independence. Consequently, the data were examined to determine if this were the case. The observation of higher inter-correlations among the several separate mathematics attitude indices than between these indices and measures of ability, achievement or general academic motivation would be one indication of independence. The observation of systematic differences or effects for the mathematics attitude indices when ability and/or general attitude factors are partialed out would also provide evidence that the indices were measuring somewhat independent factors.

A similar question at a different level is also posed for several specific attitude and interest indices (as contrasted to more general or global indices of interest) which were developed within the project to measure the more specific component attitudinal dimensions presumed to comprise a general affective reaction factor reflected in the global attitude indices. Similar criteria and observations indicated above are applicable to this question also. In addition, since these indices were developed within the project, it was possible to construct them to minimize their reflecting common factors by considering the interrelations between items to be included in the separate scales or indices.

In general, these are essentially questions of the construct validity of the various indices being used with respect to which some evidence will be presented in terms of the relations between the measures and variables observed.

II. Procedures

A. Sample

The sample consisted of 37 pairs of ninth grade algebra classes; each pair being taught by the same teacher in connection with his participation in a project investigating the achievement effects of the experimental programs. Their participation in the latter project consisted of using one of the experimental (E) programs in one of their algebra classes and their usual conventional (C) programs in a separate class and of administering designated achievement tests at the beginning and end of the school year to both classes.

The participant classes were in schools distributed over a five state area (Minnesota, Iowa, Wisconsin, North Dakota, and South Dakota). The greatest majority were in less populated communities rather than the larger metropolitan areas in this region. Participation was voluntary on the part of both teachers and school administrators.

Because the achievement evaluation project being carried out with the experimental materials had been underway for several years, teachers in the sample for this study had varying amounts of previous experience with the experimental program they were using. Also because of the voluntary nature of their participation, there were a different number of teachers using each experimental program and, therefore, the number of pairs of E and C classes, as well as the number of previous years experience the teachers had with the experimental program, varied among the experimental program conditions.

Table 1

Number of teachers following each E program and the number of previous years experience with that program.

Number of previous years	Ball State	UICSM	SMSG	Total
2	5	2	3	10
1	4	4	4	12
0	2	3	10	15
	11	9	17	37

As part of the procedure involved in participation in the achievement evaluation project, principals and teachers had been requested to assign pupils to the alternate classes on a random basis. That this was not accomplished for a few of the classes, inadvertantly or otherwise, was evident from the distribution of initial achievement test scores.[3] The classes did, however, represent a fairly wide range in level of mathematics achievement at the beginning of the year.

[3]As discussed below, these classes were not included in the analysis concerned with instructional treatment effects.

B. Experimental materials

The experimental materials used in the alternate experimental classes were those developed under the auspices of the Ball State Indiana Teachers College, The University of Illinois Committee on School Mathematics (UICSM), and the School Mathematics Study Group (SMSG). The specific ninth grade textbooks for each of these programs were respectively; Algebra I by Brumfiel, Eicholz and Shanks, Addison-Wesley, Mass. 1961; High School Mathematics Units 1-4, Revised Edition, Illinois Committee on School Mathematics, Univ. of Illinois Press, Urbana, Ill., 1962; First Course in Algebra, School Mathematics Study Group, Yale Univ. Press, New Haven, Conn., 1962. These programs are for the most part prototypes of what has been commonly characterized as "modern" mathematics.

C. Data gathering procedures

Principals and teachers who were already participating in the experimental program achievement evaluation project were contacted at the beginning of the school year and requested to cooperate in the data gathering aspects of this project. Upon indication of their willingness to do so, principals were requested to make arrangements for the questionnaires to be administered in the two mathematics classes by someone other than the teacher (preferably an administrator or counselor). Forms filled out by those administering the questionnaires indicated that they complied with this request in every instance.

The questionnaires incorporating the various measuring instruments were distributed to the schools for administration approximately 5 - 6 weeks after the beginning of the fall term. Most were administered within a week after their receipt. Revised questionnaires were again distributed for administration following the same procedure within the last two or three weeks of the spring term.

As part of their participation in the achievement evaluation project, mathematics achievement tests were administered to pupils in all classes at the beginning and end of the school year.

During the following year, schools were contacted to obtain the average grades received by pupils in the participating classes during the experimental year and the previous year.

D. Instrumentation

1. Attitude and interest indices

Data was obtained on a number of separate indices of attitudes and interests in mathematics which were based upon pupils' expressed feelings, preference, judgements and/or beliefs concerning mathematics as a school subject or as an area of activity.

The self-report indices were of two types; those previously developed outside of the present project which appeared to represent measures of a pupil's more general or global interest in mathematics and those developed within the present project to measure more specific component dimensions or factors underlying the attitudes toward or the interest in mathematics. The latter were developed to assess certain specific logically independent attitude or interest relevant factors or attitudinal components that seemed likely to be affected by variations in instructional conditions and which, consequently, might contribute differentially to the pupil's overall affective reaction or general attitude or interest with respect to mathematics.

The two previously developed measures of general interest in mathematics were the following:

The Aiken Mathematics Interest Scale (A scale), a twenty item Likert type scale in which the respondent indicates from among five alternatives, ranging from "strongly disagree" to "strongly agree", the extent of agreement with each of the statements provided concerning mathematics. The responses are logically keyed with response weights from 1 to 5 in the direction of a positive attitude toward mathematics. This scale was developed and reported by Aiken (1).

The Dutton Mathematics Attitude Scale (D scale), is a twenty-one Thurstone type scale developed and reported by Dutton (6). This scale is made up of statements representing varying degrees of positive and negative feelings, opinions or judgments about mathematics. The statements have weights determined by an a priori scaling procedure using judges familiar with the attitude or interest dimension being measured. Respondents were instructed to indicate those statements with which they most strongly agreed. Their score was the average of the weights of the items they selected.

Indices to measure the more specific factors or components of mathematics attitudes and interests were developed in the following way. Questionnaire items were constructed to obtain judgments, perceptions, feelings or reactions reflecting each of a number of attitudinal dimensions or attitude relevant instructional factors. These items were included in the questionnaire with the same response format being used for each item.

Following administration of the questionnaire, responses to each item were intercorrelated and the resulting correlation matrix factor analyzed using a principle components solution rotated to Kaiser's normal vari-max criterion. The correlations and factor analyses were used to identify items among those constructed for each of the indices that had similar factor loading patterns[4] and that would provide the highest intrascale and lowest interscale correlations. This resulted in some items being excluded from the indices for which they were constructed and some intended indices being dropped from further consideration because the items were found not to be sufficiently independent of those in other indices to warrant consideration as a separate dimension.[5] The items included in each of the resulting indices were

[4]It should be pointed out that the orthogonal factors resulting from the factor analysis were not used directly to define the attitude dimensions to be measured by the items nor was any construct interpretation of these factors attempted.

[5]One of these was an index of "perceived gain in knowledge" the items for which

those exhibiting higher intrascale item correlations and factor pattern similarity than those in alternate indices. Each indices also gave evidence of reflecting factors sufficiently independent of other indices to be considered a separate dimension.

The item analysis and index development activities outlined above were carried out for items included in both the beginning- and end-of-year questionnaires. The analys of the initial set of questionnaire items provided the basis for revision of som[illegible] and development of additional items to obtain more adequate indices from the end-of-year questionnaire.

The following are the indices developed from the questionnaire items to measure more specific attitude and interest factors.

a. Intrinsic interest - Consisting of items concerned with the degree of interest in or preference for activities involving or requiring the use of mathematics.

This scale represented an attempt to get at the aspect of interest that derives from the pupils' reaction to mathematics materials and activities as such in contrast to an interest that derives primarily from performance, competency, or general achievement motivation factors. That is, the degree of preference for math activities independent of outcomes in terms of achievement. Items in this index asked about the pupil's level of interest, and such things as how much he liked doing homework or extra reading in mathematics.

b. Perceived knowledge - Items concerned with pupils' judgement about his own knowledge or proficiency in mathematics. This index included items requesting the pupils' judgement of his own proficiency relative to other pupils as well as in absolute terms.

A pupil's conception of his own proficiency in a given subject matter area has often been suggested as a factor relevant to subsequent achievement. Recent evidence presented by Brookover (4) provides direct support for this contention.

c. Perceived utility - Items concerned with the extent to which knowledge of mathematics was seen as facilitating achievement of the pupils future goals and objectives, that is "how useful or important" they felt knowledge of the subject was for what they wanted to do later on.

This index was included in part because studies of factors underlying social attitudes have suggested that perceived instrumentality or utility of the attitude object for achieving valued goals or ends is a relevant factor influencing the intensity of the attitude. (See Rosenberg (8).) In addition, mathematics is often conceived of as a skill which is acquired primarily for practical purposes. This is a characteristic which might be less apparent in the "modern" experimental as compared to the conventional mathematics programs.

could not be distinguished from those in an index measuring "ease of learning."

d. Experienced ease or difficulty learning - Items concerned with the ease or difficulty the pupil experienced learning and understanding the material presented in the mathematics class.

Although not obviously an attitude dimension, the pupils' subjective impression of the ease or difficulty experienced in conjunction with the required learning tasks is reasonably a factor highly relevant to the pupil's affective reaction to that subject.

e. Achievement motivation - Items concerned with the pupils' desire or determination to achieve at a high level in his mathematics class, e.g. "how important is it to you to get a high grade?"

This indexwas included to tap the pupils more general motivation to achieve (in the sense suggested by Atkinson (2) as it might be reflected in his mathematics class.

In addition to the items comprising the scales providing direct measures of pupil attitudes and interests, items were also included to obtain indices of instructional factors and conditions which could possibly influence, but less directly reflect, attitudinal effects. Some of these items were incorporated into multiple item scales, others were used as single items. Among these was a multiple item index concerning the amount of homework the pupil engaged in for his mathematics class, i.e., an Expended Effort index.

Single items indices concerned with how well the pupil liked the teacher and judgements about the ease of understanding and using their texts were also included. The textbook item was included to obtain a direct pupil reaction to the experimental materials used in the classes.

2. Response format and attitude index scores

Each of the items included in the above indices was constructed to obtain a response on a graphic scale having appropriate labels accompanying the item which in effect served to define the dimension of response. Since the same response format was used for all items, the items prepared for the separate indices were included in the questionnaire as a single set of items following the same response instruction. Pupils were instructed to respond to each item for each of the academic subjects they were taking (which were designated as mathematics, English, social studies, science, and foreign language). For each item, the pupil's response for each subject was made on the same scale that accompanied that item. This form of response permitted each item to be scored for a given subject such as mathematics in two ways; (1) in terms of actual scale units for that position on the scale (absolute value or a-v) and (2) in terms of the rank position (r-p) for that subject relative to the pupils' other subjects.

This procedure was followed for several reasons. One was to eliminate certain types of response bias that might otherwise occur when responding only with respect to a single subject. One type of response bias would be the pupil's general attitude toward school which might be represented by a tendency to consistently respond toward the positive or negative end of the scale on each of the items. This tendency would seem to represent what has been characterized by some (see Rorer (7)) as a "response set." Also, scores based on relative rank responses would eliminate "response style" differences between those who tend to respond at the extremes of any scale and those who tend to respond more toward the middle of the scale, i.e. response polarization.

Another consideration was that a more objective frame of reference, which was relatively common or standard for all pupils, would be provided by having pupils respond with respect to several subjects in addition to mathematics.

In short, then for each index two scores could be obtained. One score based upon the absolute scale value of the response to each item. The other score indicating the relative position for mathematics compared to other subjects for each item.

3. Other measures

Measures of pupil achievement or proficiency in mathematics were also obtained. At the beginning and end of the school year, the mathematics section of the Sequential Tests of Educational Progress (STEP), Level Two, (10) was administered to all classes in the sample. The pupils average grades in mathematics for the previous year, eighth grade, and the concurrent year, ninth grade, were also obtained for approximately 80 percent of the classes in the sample.

One factor in the instructional situation that could affect the pupils' attitudes in addition to the materials used, was the teacher's attitude or judgment concerning the materials. Even though the teachers had volunteered to participate, it seemed likely that their judgments and evaluations of the experimental materials could vary over time. A more positive reaction to the experimental than the conventional program might be more frequently expected, reflecting in part the "Hawthorne effect" often noted in curriculum evaluation studies, but other factors could contribute to a negative evaluation of the experimental program. The nature of the teacher's reaction could carry over to the pupils, either in a direct way through the teacher's expression of his attitude or possibly indirectly in terms of overt enthusiasm, etc.

To have some assessment of the teacher's attitude and judgment about the experimental materials, a questionnaire was prepared requesting on a number of specific items, the teachers judgments, feelings, and their characterization of the experimental programs they were teaching. Thirty-five of the 37 teachers in the sample returned completed questionnaires. This questionnaire provided a basis for classifying or scoring teachers in terms of their relative attitudes and judgments about the instructional programs which could be examined for their possible correspondence with the resultant pupil attitudes.

E. Analysis

The analysis was carried out to determine whether the attitudes developed toward mathematics differed between pupils in experimental (E) and conventional (C) classes. Pupils in classes instructed with the separate experimental programs were compared with those in classes instructed by the same teachers with a conventional program. Each teacher, therefore, had an E and a C class. Since the main question concerned changes or effects occuring over the school year, it was necessary to take into account the pupils' initial level as observed at the beginning of the school year on each of the outcome variables being considered. This was done by blocking on levels of the premeasure of the dependent variable being analyzed and treating the premeasure as a separate factor in the analysis design which in effect also corrected for any initial differences between comparison groups. To make these comparisons a four-factor partially hierarchal analysis of variance design was used. The four factors were:

1. The program used in the E class - the E program comparison condition. Each teacher used one of three E programs, Ball State, UICSM and SMSG.

2. The instructional treatment - whether the class was receiving instruction with an E or a C program.

3. The premeasure control for the dependent variable - two levels determined by the median of the overall distribution of scores were used.

4. The teacher - teachers were nested within the alternate E program comparison conditions.

The instructional treatment and premeasure (or control) factors were crossed with each other and with the teacher and E program factors. A schematic representation of this design with the factors and alternate levels designated is given below.

Representation of analysis of variance design

	Treatment	B_1 (E)			B_2 (C)			B.		
	Premeasure	C_1(low)	C_2(high)	C.	C_1	C_2	C.	C_1	C_2	C.
Program	Teacher									
A_1 (Ball State)	$D_1(1)$	X_{1111}	X_{1121}	$X_{11.1}$			$X_{12.1}$	$X_{1.11}$	$X_{1.21}$	$X_{1..1}$
	$D_2(1)$	X_{1112}								
	.									
	.									
	$D_L(1)$									
	$D.(1)$	$X_{111.}$	$X_{112.}$	$X_{11..}$			$X_{12..}$	$X_{1.1.}$	$X_{1.2.}$	$X_{1...}$
A_2 (UICSM)	$D_1(2)$									
	.									
	.									
	$D_L(2)$									
	$D.(2)$	$X_{211.}$	$X_{212.}$	$X_{21..}$			$X_{22..}$	$X_{2.1.}$	$X_{2.2.}$	$X_{2...}$
A_3 (SMSG)	$D_1(3)$									
	.									
	.									
	$D_L(3)$									
	$D.(3)$	$X_{311.}$	$X_{312.}$	$X_{31..}$			$X_{32..}$			$X_{3...}$
A.		$X_{.11.}$	$X_{.12.}$	$X_{.1..}$			$X_{.2..}$			$X_{....}$

A mixed effects model, with teachers being the single random variable, was used with an unweighted means method of analysis. Use of the unweighted means solution was required because of the varying proportions of pupils in each class falling in the alternate levels or blocks for the premeasure as well as the different numbers of pupils in each class. The latter factors also necessitated adjusting the within-cells error estimate for the unequal frequencies within these cells.

The elements and notation for the four factor partially hierarchal design are shown below:

	Factor	Notation and levels		Distribution of factor levels	Sampling correction factor
A:	Program	$i = 1, \ldots I,$	where $I = 3$	fixed $I=I'$	$1-\frac{I}{I'} = 0$
B:	Treatment	$j = 1, \ldots J,$	where $J = 2$	fixed $J=J'$	$1-\frac{J}{J'} = 0$
C:	Premeasure	$k = 1, \ldots K,$	where $K = 2$	fixed $K=K'$	$1-\frac{K}{K'} = 0$
D:	Teacher	$l = 1, \ldots L,$	where L = Number of teachers	random $L << L'$	$1-\frac{L}{L'} = 1$
		$m = 1, \ldots M,$	where M = Number of pupils/cell (unequal in each cell)	random $M << M'$	$1-\frac{M}{M'} = 1$

Factor D (teacher) is nested under factor A (curriculum).

The model for this analysis has the form:

$$X_{ijklm} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \delta_{l(i)} + \beta\delta_{jl(i)}$$
$$+ \gamma\delta_{kl(i)} + \beta\gamma\delta_{jkl(i)} + \epsilon_{m(ijkl)} .$$

The degrees of freedom, expected mean squares and appropriate error terms for this kind of design are given below.

	Source of variation	Degrees of freedom	Expected mean square	Appropriate error term
1)	A	I-1	$\sigma^2_\varepsilon + JKM\sigma^2_{D(A)} + JKLM\sigma^2_A$	$MS_{D(A)}$ or $MS_{S(ABCD)}$
2)	B	J-1	$\sigma^2_\varepsilon + KM\sigma^2_{B\cdot D(A)} + IKLM\sigma^2_B$	$MS_{B\cdot D(A)}$ or $MS_{S(ABCD)}$
3)	C	K-1	$\sigma^2_\varepsilon + JM\sigma^2_{C\cdot D(A)} + IJLM\sigma^2_C$	$MS_{C\cdot D(A)}$ or $MS_{S(ABCD)}$
4)	AB	(I-1) (J-1)	$\sigma^2_\varepsilon + KM\sigma^2_{B\cdot D(A)} + KLM\sigma^2_{AB}$	$MS_{B\cdot D(A)}$ or $MS_{S(ABCD)}$
5)	AC	(I-1) (K-1)	$\sigma^2_\varepsilon + JM\sigma^2_{C\cdot D(A)} + JLM\sigma^2_{AC}$	$MS_{C\cdot D(A)}$ or $MS_{S(ABCD)}$
6)	BC	(J-1) (K-1)	$\sigma^2_\varepsilon + M\sigma^2_{BC\cdot D(A)} + ILM\sigma^2_{BC}$	$MS_{BC\cdot D(A)}$ or $MS_{S(ABCD)}$
7)	ABC	(I-1) (J-1) (K-1)	$\sigma^2_\varepsilon + M\sigma^2_{BC\cdot D(A)} + LM\sigma^2_{ABC}$	$MS_{BC\cdot D(A)}$ or $MS_{S(ABCD)}$
8)	D(A)	I(L-1)	$\sigma^2_\varepsilon + JKM\sigma^2_{D(A)}$	$MS_{S(ABCD)}$
9)	C·D(A)	(K-1)I(L-1)	$\sigma^2_\varepsilon + JM\sigma^2_{C\cdot D(A)}$	$MS_{S(ABCD)}$
10)	B·D(A)	(J-1)I(L-1)	$\sigma^2_\varepsilon + KM\sigma^2_{B\cdot D(A)}$	$MS_{S(ABCD)}$
11)	BC·D(A)	(J-1)(K-1)I(L-1)	$\sigma^2_\varepsilon + M\sigma^2_{BC\cdot D(A)}$	$MS_{S(ABCD)}$
12)	S(ABCD)	$N_{....}$ - IJKL	σ^2_ε	
Total		$N_{....}$ - 1		

In testing the effects of the nested dimension, D(A), C·D(A), B·D(A) and BC·D(A), preliminary tests are required. By using the adjusted within cells error as denominator in the F ratio, these tests were run at $\alpha = .25$. If all these tests were null (i.e. $\delta i(i)$, $\beta\delta_j e(i)$, $\gamma\delta_{ke}(i)$, and $\beta\gamma\delta_j ke(i)$ were dropped from the model) then the adjusted within cells error was used to test the remaining interactions and main effects. If all the preliminary tests were significant, then the corresponding error term shown in the table was used to test the main effects and remaining interactions.

Because the variance estimates for D(A), C·D(A), B·D(A) and BCD(A) are obtained by pooling the respective variances within each nested level (i.e. E program comparison condition), the model assumes that these variances are homogeneous. A homogeneity of variance test (F_{max} Winer (11), pp. 92-96) was used to test this assumption for each of the variance sources. Since F tests are quite robust with respect to departures from homogenity of variance the null hypothesis was rejected only when $p < .01$. When the latter hypothesis was rejected, only comparisons made within rather than across the nested condition, i.e., only an analysis for each alternate E program condition, using the model given below, could indicate the actual treatment effects.

Although there was a different number of teachers in the sample following each of the experimental programs, to assess the teacher effects, the analysis design required that an equal number of teachers be represented within each E program condition, i.e. each nested level. This meant that some selection be made among the teachers in the E program conditions having the greater number. Another condition of selection was also necessary due to the rather wide range of class differences observed for many of the premeasures which resulted in some classes having too few pupils in either the high or low levels or blocks on the premeasure to fit the minimal conditions for analysis. Both of these conditions were met by selecting in equal numbers, within each E program condition, those teachers having classes for which the cell frequencies were above the minimum necessary and which exhibited the most balanced proportions with respect to the alternate levels on the premeasure. This determination was made separately for the analysis of each of the dependent variables, i.e. each of the attitude and interest indices, since for each a different premeasure was used. Scores defining the two levels or blocks for the premeasure control variable were established by the median of the distribution of scores on this variable obtained by pupils in all classes.

Among the sources of variation in the four factor analysis design, the treatment main effects and several treatment interactions; program by treatment, treatment by premeasure and treatment by teacher were of primary interest. The treatment main effects would represent the degree to which there were E-C differences over all three E program comparison conditions. The treatment by program interaction indicates variation in the E-C differences among the alternate E program (treatment) conditions while the treatment by premeasure interaction indicates a variation in the E-C difference between pupils having higher and lower scores on the premeasure.

The program and premeasure main effects are of less interest. The program main effects would indicate the extent to which there were differences among the three E program comparison conditions considering <u>both</u> classes (E <u>and</u> C) for each teacher. A reliable program main effect would indicate that there were general differences, as reflected in both E and C classes, between the separate sets of teachers using each E program and/or that the separate E programs had some differential effect on both classes which necessarily would have to have been mediated by the teacher. That is, a program effect could be attributed to the differences among the sets of teachers following a given E program as much as to the program differences as such. The former possibility seems more plausible, however, since teacher and program effects are confounded, it would not be possible to determine which were the case. In either event they are not questions of primary importance for this study.

A reliable premeasure main effect would usually be expected since this would indicate that the differences existing at the beginning of the year on a given measure persisted through the year. The higher the pre-post correlation for a given measure, the greater should be the premeasure effect.

The teacher main effects and teacher by treatment interactions were also of somewhat less concern since they reflect general teacher differences (within E program conditions) common to both E and C classes and do not, therefore, have any clear implication for the treatment effects.

The treatment by teacher interactions were, however, given some consideration in that they would indicate differential treatment effects among teachers.

The four-factor analysis was carried out across all E program comparison conditions to determine the nature of the general instructional treatment effects common to all E programs, and to determine if there were reliable variations in the E-C differences between programs. It was of equal interest to examine the instructional treatment effects for each of the separate E programs.

For each of the E program comparison conditions, a three-factor analysis was also carried out following essentially the same analysis design as used across all E programs. For the three-factor analysis within the E program comparison conditions, the three factors were crossed and a $2 \times 2 \times t$ factorial design with t representing the number of teachers was used. Here again teachers were treated as a random variable.

The model for the within program three-factor design was:

$$X_{jklm(i)} = \mu_i + \beta_j + \gamma_k + \delta_l + \beta\gamma_{jk} + \beta\delta_{jl} + \gamma\delta_{kl} + \beta\gamma\delta_{jkl} + \epsilon_{m(jkl)(i)} .$$

The degrees of freedom, expected mean squares, and appropriate error terms are shown below:

Source of variation	Degrees of freedom	Expected mean square	Appropriate error term
B	J-1	$\sigma^2_\epsilon + KM\sigma^2_{BD} + KLM\sigma^2_B$	MS_{BD} or $MS_{S(BCD)}$
C	K-1	$\sigma^2_\epsilon + JM\sigma^2_{CD} + JLM\sigma^2_C$	MS_{CD} or $MS_{S(BCD)}$
BC	(J-1)(K-1)	$\sigma^2_\epsilon + M\sigma^2_{BCD} + LM\sigma^2_{BC}$	MS_{BCD} or $MS_{S(BCD)}$
D	L-1	$\sigma^2_\epsilon + JKM\sigma^2_D$	$MS_{S(BCD)}$
BD	(J-1)(L-1)	$\sigma^2_\epsilon + KM\sigma^2_{BD}$	$MS_{S(BCD)}$
CD	(K-1)(L-1)	$\sigma^2_\epsilon + JM\sigma^2_{CD}$	$MS_{S(BCD)}$
BCD	(J-1)(K-1)(L-1)	$\sigma^2_\epsilon + M\sigma^2_{BCD}$	$MS_{S(BCD)}$
S(BCD)	$N_{...}$ - JKL	σ^2_ϵ	
TOTAL	$N_{...}$ - 1		

The criterion for the preliminary test on the interaction effects BD, CD, BCD was $\alpha = .25$. If the null hypothesis was rejected, then MS_{BD} was used as the error term for testing the B effect, MS_{CD} for C effect, MS_{BCD} for BC effect.

The within P program condition analysis was carried out routinely independent of the results of the tests for the overall analysis. If, for the within program analysis, significant treatment main effects or treatment interactions were observed in the absence of an indication of such effects from the overall analyses, then any interpretation or generalization of the within program effect would have to be qualified since the probability of a Type I error would be increased by this practice by virtue of the dependent hypotheses.

For some variables the analysis was carried out for males and females separately as well as for the sexes combined.

Initial achievement scores obtained for the classes as well as questionnaire information from the participant teachers provided an indication that five homogeneously grouped classes (some high, some low ability) among those that would otherwise be included in the data analysis sample. Since such grouping could in itself be a source of certain differential attitudes and reactions between a pair of classes, both classes (E and C) for teachers having a homogeneously grouped class were, therefore, eliminated from the analysis directly concerned with the instructional effects. Four of these teachers were using the SMSG program and one the Ball State program.

III. Results

A. Overall E-C differences

1. General measures of interest in mathematics

a. Aiken Scale

Table 2 shows the adjusted (unweighted) means obtained for E and C class pupils in each program comparison condition on the Aiken Interest scale administered at the end of the year. Table 3 shows the results of the analysis of variance of the scores obtained on this measure across all program comparison conditions for the sexes separately and combined.

Over all programs the C class pupils had a higher mean Aiken scale score than the E class pupils which was reflected by a significant treatment main effect in the analyses. This result was obtained even though in the UICSM comparison the E class mean was higher than that for the C classes. However, the homogeneity of variance assumption for the teacher by treatment interaction term required for the treatment effects test was not tenable, i.e. there was a highly significant difference among the separate variances that were pooled to estimate this interaction effect. Consequently, only the analysis within each of the separate program comparison conditions could provide an indication of the instructional treatment effects. For these comparisons, the three-factor analysis of variance design was used, the results of which are shown in Table 4.

TABLE 2

Aiken Mathematics Interest Scale adjusted mean scores for pupils in E and C classes.

All pupils 9 pairs E and C classes in each E program condition.

	Treatment	E			C		
Experimental Program	Premeasure Level	low	high	ave.	low	high	ave.
Ball State		2.87	3.69	3.28	3.03	3.90	3.46
UICSM		2.96	3.71	3.34	2.77	3.90	3.33
SMSG		2.72	3.73	3.23	2.91	3.72	3.32
Total		2.85	3.71	3.28	2.90	3.84	3.37
Males 5 pairs E and C classes							
Ball State		3.04	3.69	3.37	2.99	3.77	3.38
UICSM		2.96	3.96	3.46	3.17	3.79	3.48
SMSG		3.06	3.72	3.39	2.85	3.93	3.39
Total		3.02	3.79	3.40	3.00	3.83	3.41
Females 7 pairs E and C classes							
Ball State		2.86	3.29	3.06	2.84	3.78	3.31
UICSM		2.93	3.79	3.36	2.56	4.11	3.33
SMSG		2.73	3.61	3.17	2.78	3.80	3.29
Total		2.84	3.56	3.20	2.72	3.89	3.31

TABLE 3

F-ratios from the analysis of variance over all E program comparison conditions for the Aiken and Dutton scale scores for males and females separately and combined.

Source of Variance	d.f.	Aiken All Pupils	Aiken M	Aiken F	Dutton All Pupils	Dutton M	Dutton F
Program	2	1.1	.4	.6	.1	.7	.9
Treatment	1	4.3*	0.0	1.4	1.0	0.0	1.1
Premeasure	1	413.4***	88.3***	172.9***	179.6***	85.1***	201.7***
Program X Treatment	2	1.4	0.0	.7	.8	.1	.9
Program X Premeasure	2	.4	.3	4.3*	3.3*	0.0	.1
Treatment X Premeasure	1	.4	.2	9.7*	1.5	.1	4.4
Program X Treatment X Premeasure	2	1.9	2.5	1.3	.2	0.0	3.3
Teacher	3(t - 1)	1.6*	1.5	2.7**	.6	2.0*	1.7*
Number of teachers	t =	9	5	7	8	7	7
Teacher X Premeasure	3(t - 1)	.9	1.3	.7	.7	1.3	.7
Teacher X Treatment	3(t - 1)	.8	.6	1.6	.4	.9	1.6*
Teacher X Treatment X Premeasure	3(t - 1)	1.8*	.9	.8	.8	1.4	1.5

* p <.05

** p <.01

*** p <.001

TABLE 4

F- ratios from the analysis of variance for the Aiken and Dutton scale scores within each of the experimental program comparison conditions.

Experimental Program	Source of Variance	d.f.	Aiken All Pupils	Aiken M	Aiken F	Dutton All Pupils	Dutton M	Dutton F
Ball State	Treatment	1	3.0	0.0	1.5	.3	0.0	1.0
	Premeasure	1	88.4***	27.5***	25.7***	105.4***	31.7***	51.5***
	Treat X Premeasure	1	0.0	.2	3.7	.5	.1	.1
	Teacher	(t - 1)[a]	1.5	1.4	2.6	1.9	2.8*	1.1
	Treat X Teacher	(t - 1)	1.8	.7	2.0	.5	1.4	1.4
	Premeas. X Teacher	(t - 1)	1.3	1.2	.7	.3	.5	.6
	Treat X Premeasure X Teacher	(t - 1)	2.8**	.9	1.2	1.9	1.8	1.8
UICSM	Treatment	1	0.0	0.0	.1	.1	.1	.2
	Premeasure	1	147.1***	40.0***	90.5***	40.4***	25.7**	75.4***
	Treat X Premeasure	1	5.7*	2.1	7.6**	3.2	.1	16.0**
	Teacher	(t - 1)	1.6	2.3	2.0	1.2	2.0	2.3*
	Treat X Teacher	(t - 1)	.2	.9	.4	.2	.7	2.1
	Premeas. X Teacher	(t - 1)	.8	.9	.2	2.5*	1.4	.8
	Treat X Premeasure X Teacher	(t - 1)	.9	1.3	.5	1.1	.7	1.0
SMSG	Treatment	1	1.6	0.0	.5	1.3	.1	3.0
	Premeasure	1	157.6***	26.9**	71.6***	51.2***	19.6**	79.4**
	Treat X Premeasure	1	1.0	3.0	.3	0.0	0.0	0.0
	Teacher	(t - 1)	1.8	.8	3.6**	.3	.9	1.9
	Treat X Teacher	(t - 1)	.5	.3	2.4*	.4	.6	1.2
	Premeas. X Teacher	(t - 1)	.4	1.9	1.3	.4	2.2*	.7
	Treat X Premeasure X Teacher	(t - 1)	1.8	.4	.7	.5	1.5	1.6

[a] t = number of teachers (i.e. pairs of classes indicated in the table of means for each measure.

* $p < .05$, ** $p < .01$, *** $p < .001$

The results of the analysis over all programs for males and females separately did not indicate a reliable treatment effect, but for the females there was a significant treatment by premeasure interaction. The latter effect resulted from the fact that among girls having initially lower interest scores, the end of year scores tended to be relatively higher for the E than the C class pupils while among girls having higher initial interests, those in the C classes had higher scores at the end of the year.

The results of the separate analysis for each program comparison condition for each sex separately are also shown in Table 4.

The analyses within each program comparison condition for all pupils indicated no significant instructional treatment main effects. The reliable treatment difference noted above when the analyses was made over all programs was apparently a spurious result of the untenable homogeneity of variance assumption. For the UICSM comparison a significant treatment by premeasure interaction was obtained. The analyses within the UICSM program comparison for males and females separately indicated that this effect (which was also observed to be significant in the analyses for females across all E treatment conditions) was highly significant for girls but non-significant for boys. A test of the E-C differences for UICSM girls within each of the initial interest levels showed that in the low interest level the E class mean was reliably higher ($F = 4.4$, $p < .05$) than that for the C classes, but that for those having initially higher interests, the difference in favor of the C class girls did not quite reach the .05 level of reliability ($F = 3.2$, $.05 < p < .10$).

b. Dutton scale

Table 5 shows the adjusted (unweighted) Dutton Attitude scale means obtained for E and C class pupils in each E program comparison condition. The analysis of variance across all program comparison conditions for both sexes separately and combined indicated no significant treatment differences (either main effects or interactions) for this measure. These results are shown in Table 3.

For the analysis within each E program condition for the Dutton scale (shown in Table 4), the only reliable treatment effect observed was a treatment by premeasure interaction for girls in the UICSM comparison. As was observed on the Aiken scale, among girls having lower interest scores at the beginning of the year, those instructed with the UICSM program had higher post instruction scores than those in the C classes while the difference was in the opposite direction for those having higher pre instruction interest scores. A test of the differences between each E-C mean within the premeasure levels indicated that for the lower level the E mean was reliably greater than the C class mean ($F = 11.1$, $p < .01$) while for the higher level, the C class mean was reliably greater ($F = 6.1$, $p < .05$).

The scales providing a more general or global measure of mathematics interest did not reveal any consistent overall differences between pupils instructed with any of the experimental programs and those instructed with conventional programs. On the Aiken scale there was a general tendency for the E-C differences to be greater in favor of the C classes among girls having higher rather than lower initial interests. This variation in the E-C difference was most pronounced and statistically reliable for girls instructed with the UICSM program. For the latter comparison on both the Aiken and Dutton scales, among girls having initially lower interests, those in UICSM classes had the higher mean interest scores while for girls with higher initial interests those in the comparison conventional classes had the higher means.

TABLE 5

Dutton Mathematics Attitude Scale adjusted mean scores for pupils in E and C classes.

All pupils 8 pairs E and C classes in each E program condition.

Experimental Program	Treatment / Premeasure Level	E low	E high	E ave.	C low	C high	C ave.
Ball State		5.10	6.60	5.85	5.02	6.85	5.94
UICSM		5.19	6.56	5.88	4.85	6.81	5.83
SMSG		4.54	6.92	5.73	4.87	7.38	6.13
Total		4.95	6.70	5.82	4.91	7.02	5.97
Males 7 pairs E and C classes							
Ball State		5.41	6.97	6.19	5.55	6.96	6.26
UICSM		5.25	6.72	5.99	5.23	6.56	5.90
SMSG		5.33	6.79	6.06	5.36	6.87	6.11
Total		5.33	6.82	6.08	5.38	6.80	6.09
Females 7 pairs E and C classes							
Ball State		4.58	6.32	5.45	4.80	6.70	5.75
UICSM		5.22	6.22	5.72	4.18	7.00	5.59
SMSG		4.34	6.09	5.21	4.67	6.44	5.56
Total		4.71	6.21	5.46	4.55	6.71	5.63

2. Indices of specific attitude dimensions

As discussed above, two scores were derived from the items comprising the indices developed to assess specific attitude and interest dimensions or factors. One score was based upon the absolute scale value response to the individual items which was designated the a-v score. The second score was based upon the response to each item for mathematics relative to other subjects, i.e. the subject rank position for mathematics which was designated the r-p score.

a. Absolute-value Scores

i. Intrinsic Interest

The Intrinsic Interest index a-v score means are shown in Table 6. The results of the analysis across all of the E program conditions are given in Table 7. In the analysis for females alone, there was a significant program by treatment interaction. This effect was a result of the UICSM girls having a higher mean and the Ball State girls having a lower mean than those in their respective conventional comparison classes. The results of the analysis for girls <u>within</u> each experimental program comparison condition shown in Table 8, indicated none of the E-C differences within the separate treatment conditions reached the .05 level of significance.

The significant interaction and the pattern of means indicates, however, that for girls the intrinsic interest of those in the UICSM program was much more positive relative to the intrinsic interest of girls in the comparison conventional classes than were the interests of girls instructed with the Ball State program compared to those in their comparison classes. No reliable treatment effects were observed in the analyses made for this measure for males or for all pupils combined.

ii. Perceived Utility

The adjusted means for scores obtained on the Perceived Utility index are shown in Table 9. The results of the analysis of variance considering all programs are shown in Table 7.

A significant program by treatment interaction was obtained for the analysis considering <u>all</u> pupils and considering females alone. This interaction in both instances appeared to be the result of the C class pupils having a higher mean Perceived Utility score in the Ball State comparison and the E class pupils having the higher mean in the UICSM comparison. The within program analysis was carried for each of the sexes separately and combined, and is shown in Table 8. The results of the latter analyses indicated that when all pupils were considered, the E-C differences for both the Ball State and UICSM comparisons were reliable at the .05 level. The analysis for girls alone indicated only the UICSM E-C difference was reliable. The latter analyses also indicated a significant treatment by prelevel interaction for the UICSM program comparison which was the result of E-C difference in favor of the UICSM classes being larger among girls having lower rather than higher initial Perceived Utility scores. For boys alone, no significant treatment effects were indicated for any of the programs.

TABLE 6

Intrinsic Interest adjusted mean index scores for pupils in E and C classes.

All pupils 9 pairs E and C classes in each E program condition.

	Treatment	E			C		
Experimental Program	Premeasure Level	low	high	ave.	low	high	ave.
Ball State		2.47	3.44	2.96	2.48	3.38	2.93
UICSM		2.55	3.27	2.91	2.44	3.34	2.89
SMSG		2.46	3.43	2.94	2.47	3.51	2.99
Total		2.49	3.38	2.94	2.46	3.41	2.94
Males 6 pairs E and C classes							
Ball State		2.23	3.50	2.86	2.40	3.45	2.92
UICSM		2.57	3.36	2.97	2.43	3.26	2.86
SMSG		2.54	3.50	3.02	2.34	3.74	3.04
Total		2.45	3.45	2.95	2.39	3.48	2.94
Females 6 pairs E and C classes							
Ball State		2.49	2.94	2.71	2.76	3.35	3.05
UICSM		3.05	3.37	3.21	2.47	2.95	2.71
SMSG		2.46	3.10	2.78	2.37	3.11	2.74
Total		2.67	3.13	2.90	2.53	3.13	2.83

TABLE 7

F-ratios obtained from the analysis of variance over all E program comparison conditions for each attitude index s-v score for males and females separately and combined.

Source of Variance	d.f.	Intrinsic Interest All pupils	Intrinsic Interest M	Intrinsic Interest F	Perceived Knowledge All pupils	Perceived Knowledge M	Perceived Knowledge F	Perceived Utility All pupils	Perceived Utility M	Perceived Utility F	Ease of Learning All pupils	Ease of Learning M	Ease of Learning F
Program	2	.2	.4	.5	1.2	1.2	1.5	.4	.4	.3	.8	2.8	.9
Treatment	1	0.0	0.0	.3	3.2	1.4	0.0	1.2	.8	.6	10.5**	1.3	14.2**
Premeasure	1	217.7***	80.3***	17.4***	243.7***	76.1***	14.0***	74.4***	35.3***	43.3***	197.2***	72.6***	100.3***
Program X Treatment	2	.1	.3	3.9*	.8	1.1	1.9	5.4**	.7	3.7*	.3	.2	1.6
Program X Premeasure	2	.8	1.0	.4	1.8	.6	0.0	.6	5.4**	0.0	1.4	0.0	3.0
Treatment X Premeasure	1	.2	.1	.4	0.0	.1	0.0	.3	.2	2.2	0.0	1.0	1.0
Program X Treatment X Premeasure	2	.3	.6	0.0	1.7	.5	1.2	1.1	.4	1.4	1.0	.7	.2
Teacher	3(t - 1)	2.3***	2.0*	2.7***	2.5***	.7	2.4***	3.4***	1.3	3.1***	2.3***	.7	2.[illegible]
Number of teachers	t =	9	6	6	8	6	6	8	5	6	8	6	5
Teacher X Premeasure	3(t - 1)	1.1	1.4	1.6	1.0	1.1	.6	1.4	1.1	.8	1.3	1.0	1.3
Teacher X Treatment	3(t - 1)	.9	.7	1.4	1.0	1.1	1.0	.9	.6	.9	1.0	1.2	1.6
Teacher X Treatment X Premeasure	3(t - 1)	.9	1.6	1.2	1.3	1.5	1.1	1.5	1.4	.9	.9	.4	1.1

* p <.05, ** p <.01, *** p <.001

TABLE 8

F-ratios obtained from the analysis of variance within each of the E program comparison conditions for the Intrinsic Interest and Perceived Utility index s-v scores.

Experimental Program	Source of Variance	d.f.	Intrinsic Interest: All Pupils	Intrinsic Interest: M	Intrinsic Interest: F	Perceived Utility: All Pupils	Perceived Utility: M	Perceived Utility: F
	Treatment	1	.1	.1	2.7	2.2*	0.0	1.3
	Premeasure	1	67.2**	40.6***	6.6*	49.1***	29.1***	12.5***
	Treat X Premeasure	1	.1	.4	.1	.4	0.0	.1
Ball State	Teacher	(t - 1)[a]	3.1***	2.4*	4.4**	1.2	.6	1.0
	Treat X Teacher	(t - 1)	.8	1.0	.6	1.9	1.0	.9
	Premeas. X Teacher	(t - 1)	.7	1.1	.9	1.2	.1	.8
	Treat X Premeasure X Teacher	(t - 1)	.3	.5	1.0	1.9	1.5	1.3
	Treatment	1	0.0	.5	4.0	4.4*	2.6	6.2*
	Premeasure	1	57.5***	11.9*	1.8	19.9***	3.2	14.3***
	Treat X Premeasure	1	.5	0.0	.1	1.0	.1	4.0*
UICSM	Teacher	(t - 1)	1.8	2.5*	1.3	5.2**	2.4	5.0***
	Treat X Teacher	(t - 1)	1.0	.7	2.2	.4	.3	.7
	Premeas. X Teacher	(t - 1)	1.2	1.8	3.2*	1.6	1.8	.4
	Treat X Premeasure X Teacher	(t - 1)	1.4	1.8	2.1	1.5	1.4	1.0
	Treatment	1	.2	0.0	0.0	3.5	0.0	0.0
	Premeasure	1	62.3***	52.7***	16.0***	18.4***	2.3	16.9***
	Treat X Premeasure	1	.2	.7	.1	1.9	1.4	.6
SMSG	Teacher	(t - 1)	1.8	.7	1.4	3.8***	1.1	3.4***
	Treat X Teacher	(t - 1)	1.0	.3	1.6	.6	.1	1.1
	Premeas. X Teacher	(t - 1)	1.6	1.2	.7	1.5	1.8	1.3
	Treat X Premeasure X Teacher	(t - 1)	1.2	2.9*	.6	1.1	1.1	2.0

[a] t = number of teachers (i.e. pairs of classes) indicated in the table of means for each measure.

* $p < .05$, ** $p < .01$, *** $p < .001$

TABLE 9

Perceived Utility adjusted mean index scores for pupils in E and C classes.

All pupils 8 pairs E and C classes in each E program condition.

Experimental Program	Treatment / Premeasure Level	E low	E high	E ave.	C low	C high	C ave.
Ball State		4.21	5.21	4.71	4.58	5.49	5.04
UICSM		4.66	5.31	4.98	4.26	5.06	4.66
SMSG		4.13	5.06	4.60	4.50	5.17	4.83
Total		4.33	5.19	4.76	4.45	5.24	4.84
Males 5 pairs E and C classes							
Ball State		4.29	5.58	4.94	4.27	5.64	4.95
UICSM		4.97	5.55	5.26	4.73	5.13	4.93
SMSG		5.00	5.17	5.08	4.72	5.38	5.05
Total		4.75	5.43	5.09	4.57	5.38	4.98
Females 6 pairs E and C classes							
Ball State		3.78	4.72	4.25	4.12	4.94	4.53
UICSM		4.37	4.81	4.59	3.23	4.69	3.96
SMSG		3.75	4.51	4.13	3.60	4.71	4.15
Total		3.96	4.68	4.32	3.65	4.78	4.21

In general these results indicate that pupils instructed with the Ball State materials had a tendency to perceive mathematics as having less utility than did pupils in conventional classes taught by the same teachers while pupils instructed with the UICSM materials perceived mathematics as having greater utility than did the pupils in the comparison conventional classes. This effect appeared to be stronger for girls than boys. Also for girls, the E-C difference for the UICSM program was more reliable than the E-C difference for the Ball State comparison.

It appears then, that instruction with the Ball State program resulted in a perception of mathematics as having less utility for future goals or objectives while instruction with the UICSM program resulted in a perception of greater utility.

iii. Perceived Knowledge

The adjusted mean Perceived Knowledge index scores for the E and C class pupils are shown in Table 10. No significant treatment main effects or treatment interaction effects were indicated by the analysis of variance when all program comparison conditions were considered. The results of this analysis are shown in Table 7.

For the analysis on this measure within each E program condition, shown in Table 11, a reliable treatment difference was indicated for the SMSG comparison when all pupils and males alone were considered. This effect was a result of the SMSG pupils obtaining lower perceived knowledge scores at the end of the year than conventional class pupils. The effect for all pupils combined appears to be due primarily to differences in this regard for boys rather than girls.

In general the experimental programs do not appear to have any extensive effects on pupils' judgments of their knowledge of mathematics as measured by this index. There was a tendency for boys instructed with the SMSG program to judge their knowledge somewhat lower than did boys in the conventional comparison classes.

iv. Ease of Learning

Table 12 shows the adjusted mean scores obtained by E and C class pupils on the Ease of Learning (EOL) index. In each program comparison condition, the means for pupils in the C classes were in every instance higher than those for pupils in the respective E classes. That is, the C class pupils reported greater ease of learning (i.e., less difficulty learning) the subject matter in their mathematics class than E class pupils. The results of the overall analysis of variance, shown in Table 7, indicate that the E-C difference over all program comparison conditions is quite reliable both when all pupils and when girls alone are considered. The analysis for boys alone did not indicate any significant instructional treatment differences suggesting that the effect observed with the sexes combined is due more to the differences in this regard for girls than for boys.

The results of an analysis within each of the programs for both sexes separately and combined are shown in Table 11. Considering all pupils the E-C difference was reliable at the .05 level only for the UICSM program comparison, while for girls a highly reliable difference was observed for the Ball State program comparison. No treatment differences were observed for boys for any of the comparisons.

TABLE 10

Perceived Knowledge adjusted mean index scores for pupils in E and C classes.

All pupils 8 pairs E and C classes in each E program condition.

Experimental Program	Treatment / Premeasure Level	E low	E high	E ave.	C low	C high	C ave.
Ball State		8.09	10.47	9.28	8.72	10.46	9.59
UICSM		8.60	10.07	9.33	8.26	10.42	9.34
SMSG		7.56	10.12	8.84	8.10	10.42	9.26
Total		8.08	10.22	9.15	8.36	10.43	9.39
Males 6 pairs E and C classes							
Ball State		8.49	10.00	9.25	7.98	10.37	9.18
UICSM		8.47	10.29	9.38	8.55	10.52	9.53
SMSG		7.90	10.69	9.30	8.95	11.17	10.06
Total		8.29	10.33	9.31	8.49	10.69	9.59
Females 6 pairs E and C classes							
Ball State		8.05	10.50	9.28	8.80	10.50	9.65
UICSM		8.80	10.62	9.71	7.86	10.48	9.17
SMSG		7.74	9.89	8.81	7.90	10.02	8.96
Total		8.19	10.34	9.26	8.18	10.33	9.26

TABLE 11

F-ratios obtained from the analysis of variance within each of the E program comparison conditions for Perceived Knowledge and Ease of Learning index s-v scores.

			Perceived Knowledge			Ease of Learning		
Experimental Program	Source of Variance	d.f.	All Pupils	M	F	All Pupils	M	F
Ball State	Treatment	1	1.0	0.0	1.1	3.1	.9	20.4***
	Premeasure	1	81.0***	23.3***	35.0***	56.7***	22.9***	14.0*
	Treat X Premeasure	1	.9	.5	1.1	1.2	2.5	0.0
	Teacher	(t - 1)[a]	2.2*	1.0	1.1	1.4	1.1	1.2
	Treat X Teacher	(t - 1)	1.7	2.4*	.6	.9	1.1	.8
	Premeas. X Teacher	(t - 1)	.9	.4	.5	1.2	1.1	1.7
	Treat X Premeasure X Teacher	(t - 1)	2.2*	2.5*	.9	1.0	.4	1.3
UICSM	Treatment	1	0.0	.1	1.9	4.9*	.4	2.2
	Premeasure	1	48.9***	16.5***	32.5***	86.0***	23.0***	71.9***
	Treat X Premeasure	1	1.7	0.0	1.1	.2	0.0	.6
	Teacher	(t - 1)	.9	.6	1.8	1.3	.5	1.5
	Treat X Teacher	(t - 1)	.7	.7	.8	.5	1.9	.9
	Premeas. X Teacher	(t - 1)	.4	.9	.3	1.0	.3	.4
	Treat X Premeasure X Teacher	(t - 1)	.7	1.2	1.2	.4	.4	.2
SMSG	Treatment	1	3.9*	4.0*	.6	1.4	.1	1.6
	Premeasure	1	70.6***	17.9**	48.1***	65.2***	13.7*	21.5***
	Treat X Premeasure	1	.3	.6	0.0	.5	0.0	.5
	Teacher	(t - 1)	5.1**	2.3*	5.0**	4.8***	.5	5.1***
	Treat X Teacher	(t - 1)	.5	.3	1.6	1.7	.4	3.1
	Premeas. X Teacher	(t - 1)	1.9	2.4*	1.0	1.8	2.0	1.9
	Treat X Premeasure X Teacher	(t - 1)	1.1	.9	1.2	1.4	.5	1.8

[a] t = number of teachers (i.e. pairs of classes) indicated in the table of means for each measure.

* $p < .05$, ** $p < .01$, *** $p < .001$

TABLE 12

Ease of Learning adjusted mean index scores for pupils in E and C classes.

All pupils 8 pairs E and C classes in each E program condition.

Treatment		E			C		
Experimental Program	Premeasure Level	low	high	ave.	low	high	ave.
Ball State		2.57	3.66	3.12	2.93	3.74	3.34
UICSM		2.55	3.74	3.14	2.79	4.10	3.44
SMSG		2.49	3.60	3.04	2.57	3.86	3.21
Total		2.54	3.66	3.10	2.76	3.90	3.33
Males 6 pairs E and C classes							
Ball State		2.26	3.60	2.93	2.79	3.47	3.13
UICSM		2.70	3.84	3.27	2.94	3.99	3.46
SMSG		2.74	3.82	3.28	2.79	3.87	3.33
Total		2.57	3.76	3.16	2.84	3.77	3.30
Females 5 pairs E and C classes							
Ball State		2.63	3.60	3.12	3.53	4.52	4.03
UICSM		2.32	3.92	3.12	2.47	4.39	3.43
SMSG		2.68	3.75	3.21	2.93	4.35	3.64
Total		2.54	3.76	3.15	2.98	4.42	3.70

In general, although the effect was not very large there was a consistent tendency for pupils instructed with each of the experimental programs to report more difficulty learning mathematics than pupils in the comparison conventional classes. This tendency was more evident for girls than boys and most pronounced for girls instructed with the Ball State program.

b. Rank Position scores

Indices measuring specific attitude and interest dimensions were derived from item responses in terms of the rank position for mathematics relative to other subjects. Since pupils differed as to both number and actual academic subjects other than mathematics in which they were enrolled, the mathematics ranks were determined with respect to the combination of academic subjects which would permit inclusion of the largest number of pupils for comparison. Within the sample, among the possible 3, 4, and 5 subject combinations of mathematics, English, science, social studies and foreign language, the largest number of pupils were enrolled in a 3 subject combination of mathematics, English and science.

Scores in terms of the rank position of mathematics relative to English and science were derived by summing the rank of the response for math (1, 2 or 3) to each item in a given index. Individual scores for each index were then converted to standard scores having a mean of 50 and a standard deviation of 10.

The determination of mathematics rank position relative to two other specified subjects did reduce somewhat the number of pupils that could be included in the analysis, i.e., only those taking English and science in addition to mathematics. Although the overall proportion reduced was relatively small for the r-p scores, this additional restriction did preclude carrying out a separate analyses for each sex when teachers were treated as a separate dimension because of the increased difficulty of obtaining sufficient frequencies in the separate cells required for the analysis. For these scores the sex differences will be examined in connection with other analyses of the data.

For the r-p scores in addition to the analysis of variance across all E program comparison conditions, an analysis within each E treatment condition was also carried out.

Table 13 shows the unweighted (adjusted) means obtained by E and C class pupils in the separate instructional treatment conditions on each of the indices for which rank position scores were determined. The results of the comparisons among these means considering all of the E programs for each indices are shown in terms of the F-ratios obtained from each analysis of variance in Table 14.

i. Intrinsic Interest

No significant treatment main effects nor program by treatment interactions were observed for the index measuring Intrinsic Interest using r-p scores in the analysis across all E program conditions as shown in Table 12. The higher mean for the C class pupils in the Ball State comparison contributed to a significant treatment effect when the analysis was made within the separate program conditions shown in Table 15.

TABLE 13

Adjusted means for E and C class pupils on indices using rank-position scores.

		E			C		
Premeasure Level		low	high	ave.	low	high	ave.
Index	E program						
Intrinsic Interest	BSP	44.4	49.7	47.1	46.3	52.8	49.5
	UICSM	47.6	53.4	50.5	46.8	52.2	49.5
	SMSG	46.2	55.0	50.6	47.0	55.7	51.3
	TOTAL	46.1	52.9	49.4	50.5	53.5	50.1
Perceived Utility	BSP	45.6	50.4	48.0	48.1	52.4	50.2
	UICSM	48.2	54.0	51.1	44.3	53.4	48.8
	SMSG	45.7	53.0	49.3	48.9	56.9	52.9
	TOTAL	46.5	52.5	49.5	47.1	54.3	50.7
Perceived Knowledge	BSP	43.5	52.7	48.1	47.6	54.7	51.2
	UICSM	48.4	51.6	50.0	45.3	53.9	49.6
	SMSG	44.2	53.1	48.7	45.5	52.9	49.2
	TOTAL	45.4	52.5	48.9	46.1	53.8	50.0
Ease of Learning	BSP	43.7	51.5	47.6	48.9	54.7	51.8
	UICSM	45.7	52.3	49.0	45.5	55.1	50.3
	SMSG	43.1	50.6	46.8	46.0	54.0	50.0
	TOTAL	44.2	51.5	47.8	46.8	54.6	50.7

TABLE 14

F-ratios from the analysis of variance over all E program comparison conditions for each attitude index r-p score.

Scales	d.f.	Intrinsic Interest	Utility	Ease of Learning	Perceived Knowledge
Program	2	2.0	2.8	0.0	.4
Treatment	1	1.5	2.0	11.5**	3.0
Premeasure	1	92.0***	125.7***	99.4***	79.2***
Program X Treat	2	2.6	4.5*	1.0	2.8
Program X Premeasure	2	2.1	2.9	.6	.8
Treat X Premeasure	1	0.0	1.0	.2	.2
Program X Treat X Premeasure	2	.2	.9	.8	3.3
Teacher	3(t - 1)	3.2***	1.5	3.0***	1.9*
Number of teachers	t =	8	8	7	8
Teacher X Premeasure	3(t - 1)	1.3	.8	1.6*	1.9*
Teacher X Treat	3(t - 1)	.9	2.0**	2.5**	.9
Teacher X Treat X Premeasure	3(t - 1)	.9	.8	1.3	1.2

* $p < .05$

** $p < .01$

*** $p < .001$

TABLE 15

F-ratios from the analysis of variance of each attitude index r-p score for each E program comparison condition.

Program	Source of Variance	d.f.	Intrinsic Interest	Perceived Knowledge	Perceived Utility	Ease of Learning
	Treatment	1	5.0*	8.2**	2.0	8.0*
	Premeasure	1	27.9***	59.4***	19.4***	18.9**
	Teacher	(t - 1)[a]	1.5	1.3	1.0	4.3
Ball State	Treat X Premeasure	1	.3	.5	.1	.9
	Treat X Teacher	(t - 1)	1.0	.8	2.4*	2.1
	Premeasure X Teacher	(t - 1)	1.0	.6	.5	2.4*
	Treat X Premeasure X Teacher	(t - 1)	.8	2.1	.5	.6
	Treatment	1	.8	.1	4.5*	1.0
	Premeasure	1	26.6***	26.1	48.5***	52.0***
	Teacher	(t - 1)	1.6	1.1	1.5	1.0
UICSM	Treat X Premeasure	1	0.0	5.8*	2.4	1.1
	Treat X Teacher	(t - 1)	.4	.5	.9	1.4
	Premeasure X Teacher	(t - 1)	.6	1.0	.7	1.2
	Treat X Premeasure X Teacher	(t - 1)	1.0	1.0	.7	1.6
	Treatment	1	.4	.2	4.8	1.8
	Premeasure	1	31.2***	15.8**	67.1	29.3***
	Teacher	(t - 1)	7.1***	4.0***	2.0	8.3***
SMSG	Treat X Premeasure	1	0.0	.7	.1	.1
	Treat X Teacher	(t - 1)	1.3	1.7	3.1**	6.0***
	Premeasure X Teacher	(t - 1)	2.4*	4.7***	1.3	2.3*
	Treat X Premeasure X Teacher	(t - 1)	.9	.4	1.4	1.4

[a] t = number of teachers (i.e. pairs of classes) indicated in the table of means for each measure.

* $p < .05$, ** $p < .01$, *** $p < .001$

ii. Perceived Utility

For the index of Perceived Utility, the overall instructional treatment differences were not reliable, however, a significant program by treatment interaction was obtained. As can be seen in Table 13, this resulted from the E class pupils instructed with the Ball State and SMSG programs having lower scores than those in the C classes with which they were compared while the UICSM pupils had a higher mean score than those in their comparison classes. The reliability of the E-C differences for the separate program comparison conditions were examined further with the three factor (instructional treatment by premeasure by teacher) analysis of variance design. The results of these analysis, which are shown in Table 15, indicate that only for the UICSM program was the treatment effect significant. It appears that among the E programs, only the pupils instructed with the UICSM materials exhibited a greater tendency than those instructed with the conventional materials to develop a perception of mathematics as having relatively more utility than other subjects.

iii. Perceived Knowledge

Some variation in the direction of the E-C differences for the separate experimental programs can be seen in Table 13 for the index of Perceived Knowledge. Although neither the treatment main effects nor the program by treatment interactions reached the .05 level of significance in the analysis considering all program comparison conditions shown in Table 14, they were all just beyond the .05 level, i.e. $p < .10$.

The results of the analysis for the separate E programs are shown in Table 15. Two significant treatment effects were observed. Pupils instructed with the Ball State program had reliably lower perceived knowledge scores than those in the comparison C classes. For the UICSM program a significant treatment by premeasure interaction was obtained. This resulted from the fact that among pupils at the lower level on the premeasure those in the UICSM classes had the higher mean while among pupils at the higher level on the premeasure, those in the C classes had a higher mean at the end of the year. A further test to determine if the mean differences within each premeasure level differed from zero showed that neither of these differences was highly reliable. For the lower premeasure level, $E > C$, $F = 3.8$, $.05 < p < .10$ and for the higher premeasure level, $C > E$, $F = 2.1$, $.10 < p < .25$. Nonetheless, the interaction indicates that there was a greater relative gain in perceived knowledge for UICSM instructed pupils who initially perceived their knowledge as relatively low than for those who had initially perceived their knowledge as relatively high.

These results in general indicate that at the end of the year pupils in the Ball State program tended more than those in the conventional program to perceive their knowledge of mathematics as being lower relative to their knowledge in other subjects.

However, for pupils in the UICSM program, those that had lower perceived knowledge at the beginning of the year developed a perception of relatively greater knowledge in mathematics than did similar pupils in the C classes.

iv. Ease of Learning

The analysis across all E program comparison conditions (Table 14) showed a highly reliable treatment difference with pupils in the C classes having a higher mean score as shown in Table 13. This indicates that over all program comparison conditions, C class pupils reported greater learning ease for mathematics relative to other subjects than did pupils in the E classes. Considering the separate programs, the E-C difference for the Ball State comparison was largest while that for the UICSM comparison was smallest. The analysis within each E program condition, shown in Table 15, revealed that the difference for the Ball State program was quite reliable, while the differences for the other program comparisons did not reach the .05 level of significance.

Reviewing the results obtained from the analysis concerned with the effects of the experimental programs on the several dimensions of pupil attitude toward mathematics, it appears that they were quite similar for the two types of scores used. For both the r-p and a-v response scores, a consistently lower ease of learning score was obtained for pupils instructed with each of the E programs, an effect which was most pronounced for the Ball State program. For the index of intrinsic interest when the r-p scores were used the results indicated a reliable tendency for the Ball State pupils to have lower intrinsic interest scores than did the conventional class pupils. A similar but nonreliable trend was observed for the a-v scores obtained from this scale. No reliable E-C differences were obtained for either score on this scale for the UICSM and SMSG program comparisons.

On the index of perceived utility, when either r-p or a-v scores were used, the direction of the E-C difference varied significantly between the UICSM program and the Ball State program. Both scores for UICSM instructed pupils were reliably higher than those for pupils in the comparison conventional classes. However, only for the a-v score was the difference which favored the C classes in the Ball State comparison statistically reliable.

On the index of perceived knowledge there was some variation in significant differences indicated depending on whether the r-p or a-v scores were used. For the a-v scores a treatment difference was observed only when the sexes were considered separately. The SMSG instructed boys had reliably lower a-v scores than those in the C classes. This was the only instructional treatment difference observed on these measures for the SMSG program comparison. When r-p scores were used, pupils in the Ball State program had lower scores than their C class counterparts. Also UICSM pupils having lower premeasure scores showed a higher perceived knowledge score at the end of the year than did similar C class pupils in contrast to the E-C difference for those having higher premeasure scores which favored the C class pupils.

It should be noted, however, that most of these differences are not very large, in most instances accounting for a considerably smaller proportion of the variance than is accounted for by the premeasure of each of the variables.

Considering the results of the analyses to determine the direct effects of the experimental programs on indices of both general and specific attitudes and interests in mathematics, several more general observations seem warranted. (1) The experimental programs appear to have a relatively small effect, either positive or negative, on the attitudes and interests pupils develop in the ninth grade at least as indicated by the most direct indices of these attitudes used in this study. (2) The Ball State program appeared to have a more negative than positive effect on the attitudes pupils develop toward mathematics than did comparison conventional programs of instruction. This effect was most evident for the index of perceived knowledge. (3) The UICSM program was the only experimental program for which pupils exhibited a tendency to develop more positive attitudes toward mathematics than pupils in the comparison conventional classes and even for this program, these effects were quite limited. The largest effect for the UICSM program was on the perceived utility index. (4) For all experimental programs, there was a consistent tendency for pupils instructed with the experimental materials to experience more learning difficulty than was reported by pupils in the conventional classes.

B. Specific Instructional Factors Contribution to the Experimental - Conventional Program Differences.

The above analysis has been concerned primarily with the overall effects of the different experimental programs or materials as indicated by comparative changes in pupil's attitudes toward mathematics over the year. Some differences in the resultant attitudes toward mathematics were observed. These differential outcomes were however necessarily determined or mediated by any of a number of more specific factors or conditions which must have differed among the alternate treatment conditions. Among these possible factors there were those that represented the distinctive characteristics or qualities of the different instructional materials as well as less relevant concomitant factors that may have had only incidental but nonetheless systematic relations with the program differences. The next question to be examined then concerns the factors or conditions in the instructional situation which might have contributed to the differences obtained.

With respect to this question, there were two main sets of factors or conditions that required consideration. Since they directly reflect or represent the major instructional difference, qualities or characteristics associated with the instructional materials themselves are the most obvious and likely source of any observed E-C attitude differences. However since other factors could also be involved such effects need to be demonstrated in terms of direct pupil reactions to qualities of the materials. A second set of factors are those associated with or directly affected by the teacher as such. Grading practices, instructional approach or methods, effort demands such as homework, expressed attitudes toward the materials, all are conditions which may have varied in a systematic way between the E-C classes and thereby contributed to the observed differences.

Because they involve factors most directly related to the major focus of the study pupil judgments or reactions to their instructional materials will be considered first in some detail. The teacher connected factors will be discussed in more detail below.

There were two questions to be considered with respect to pupil judgments concerning the instructional materials (1) whether the judgments varied for pupils in the E and C classes and (2) whether such judgment differences might have contributed to or could account for any of the more general attitude differences observed between E and C class pupils.

Pupil judgments were obtained concerning the degree of difficulty using and understanding the textbooks with which they had previously been instructed.

Textbook difficulty

Judgments of textbook difficulty were of interest not only for the general reasons indicated above but also because the most distinctive E-C difference was obtained for a more general index of ease of learning which presumably reflected all sources of learning difficulty including that associated with the textbook as such.

With respect to the latter index there is, then, the more specific question of whether the tendency to perceive more difficulty in general on the part of the E class pupils was a result of factors associated with the instructional materials or the teacher or both. Judgments concerning text difficulty would be considered a possible factor in this effect for both empirical as well as logical reasons since the single item index of textbook difficulty[6] was one of the items comprising the more general Ease of Learning (EOL) index.

The E-C differences in judgment of textbook difficulty were examined using the rank position response given by pupils to this item for mathematics relative to English and science. The greater the rank position value, i.e. 3, the relatively greater the difficulty. To make an appropriate comparison, it was necessary to take into account the pupils initial (beginning of year) judgment of learning ease or difficulty for mathematics materials to ensure that any obtained differences reflected the pupils experience during the year rather than his previous experience with mathematics materials. Since no textbook difficulty judgment was obtained at the beginning of the year the rank position score on the premeasure of the EOL index was used for this purpose. Comparisons were made within each E program comparison condition for pupils above and below the EOL premeasure median for each of the sexes separately. The math rank position frequencies are shown in Table 16.

It can be seen that pupils instructed with the Ball State and SMSG programs indicated much more frequently than pupils instructed with the respective conventional programs that their math textbooks were more difficult than their English or science textbooks. For the UICSM comparison the difference was not quite so evident.

To determine if the E-C frequency differences were statistically reliable in one or the other direction, a χ^2 test was used with the 1st and 2nd rank position frequencies combined.[7] To obtain at the same time an indication of the relative magnitude of the affects associated with initial or expected ease of learning (pre EOL) and with pupil sex as well as with the instructional treatment, a procedure outlined by Castellan (12) was followed which permitted determination of the relative contribution to an overall χ^2 of each of these factors.

[6]This item read, "how easy or difficult did you find it to understand the textbooks used in each of the subjects you have been taking?"

[7]The two adjacent rank position categories could be combined and not alter the interpretation because the categories had an ordered relationship. Reducing to two

TABLE 16

Text difficulty rank position frequencies

Sex	t.d. rank	Pre EOL Level	Ball State E low	Ball State E high	Ball State E total	Ball State C low	Ball State C high	Ball State C total	UICSM E low	UICSM E high	UICSM E total	UICSM C low	UICSM C high	UICSM C total	SMSG E low	SMSG E high	SMSG E total	SMSG C low	SMSG C high	SMSG C total
M	1		0	10	10	5	20	25	4	15	19	5	13	18	1	6	7	5	21	26
	2		2	12	14	8	27	35	7	11	18	6	16	22	5	9	14	14	17	31
	3		29	22	51	24	18	42	19	20	39	18	15	33	27	41	68	21	20	41
			31	44	75	37	65	102	30	46	76	29	44	73	33	56	89	40	58	98
F	1		4	5	9	5	12	17	0	7	7	6	6	12	2	5	7	10	17	27
	2		16	8	24	16	18	34	7	9	16	12	12	24	6	8	14	14	17	31
	3		49	24	73	21	9	30	30	11	41	33	8	41	62	23	85	26	9	35
			69	37	106	42	39	81	37	27	64	51	26	77	70	36	106	50	43	93
All Pupils	1		4	15	19	10	32	42	4	22	26	11	19	30	3	11	14	15	38	53
	2		18	20	38	24	45	69	14	20	34	18	28	46	11	17	28	28	34	62
	3		78	46	124	45	27	72	49	31	80	51	23	74	89	64	153	47	29	76
			100	81	181	79	104	183	67	73	140	80	70	150	103	92	195	90	101	191

The frequencies for and results of this analyses are shown in Table 17.

For both the Ball State and SMSG programs the E-C comparisons within each sex by pre EOL category show a reliable ($p < .05$) difference in frequency (the value for χ^2abc) with a greater proportion of E class pupils in each instance ranking their math text as more difficult. For the UICSM program none of the within category χ^2's (χ^2abc) are reliable at the .05 level although there is a tendency for E class pupils to more frequently rank their math texts lower (i.e. as more difficult). No reliable sex differences (χ^2ab with $p < .05$) in text difficulty judgments appeared for any of the programs, however for the Ball State comparison among pupils expecting more learning difficulty (low EOL) there was a greater tendency for boys to indicate more text difficulty than girls. With respect to levels of initial or expected learning ease (pre EOL) reliable differences (χ^2a) were observed for all programs. These differences indicate that pupils expecting more learning difficulty in general for math (low EOL) more frequently reported greater text difficulty than those in the high EOL category which was a reflection of initial pupil characteristics affecting both indices, as would be expected.[8]

It is quite evident that the Ball State and SMSG instructed pupils' assessment of their materials as being relatively more difficult was reliably different from that of the conventional class pupils with whom they were compared. For the UICSM pupils there were no real differences in this regard although there also was no tendency for the UICSM pupils to judge their materials as less difficult.

The next question to consider is whether the E-C differences observed for the EOL index were mainly a result of the textbook difficulty judgments or whether there were other factors in the instructional situation contributing to a perception of greater learning difficulty in general for pupils in the E classes.

To examine the differences between E and C classes with respect to ease of learning with adjustments for differences in judgments or perceptions of textbook difficulty, EOL means were compared within textbook judgment levels or categories using analysis of variance. A variation of the analysis design used to test the instructional treatment effects for the various attitude indices presented above was followed. For this analysis, two additional dimensions were included, sex and text difficulty with pupils pooled across teachers within each program condition. Since the E-C differences for text difficulty judgments varied among the different E programs, the analysis was made within each E program comparison condition. An unweighted means solution was again used. The adjusted EOL means obtained when sex and the text difficulty judgments were considered are shown in Table 18[9].

---

response categories permitted an unambiguous interpretation of the direction of the significant difference between E and C classes.

8Because of the procedures used to compute the χ^2 values, the differences associated with pre EOL indicated by χ^2a were independent of the sex and text difficulty affects.

9With one exception, a difficult to interpret third order interaction, there were no reliable first or higher order sex by treatment interactions. Since these were the only sex effects of concern, the tables show the adjusted means for the sexes combined, i.e. averaged over both sexes.

TABLE 17

Text difficulty rank frequency comparisons between E and C conditions considering sex and pre EOL level.

	t.d. rank	Male (b_1) $E(c_1)$	Male (b_1) $C(c_2)$	Male (b_1) Tot	Female (b_2) $E(c_1)$	Female (b_2) $C(c_2)$	Female (b_2) Tot	Total
Ball State								
Low Pre EOL (a_1)	1,2	2	13	15	20	21	41	56
	3	29	24	53	49	21	70	123
		31	37	68	69	42	111	179
		$\chi^2 abc = 6.5$*			$\chi^2 abc = 4.1$*			$\chi^2 ab = 3.7$
High Pre EOL (a_2)	1,2	22	47	69	13	30	43	112
	3	22	18	40	24	9	33	73
		44	65	109	37	39	76	185
		$\chi^2 abc = 4.7$*			$\chi^2 abc = 11.9$***			$\chi^2 ab = .6$
								$\chi^2 a = 6.9$**
UICSM								
Low Pre EOL (a_1)	1,2	11	11	22	7	18	25	47
	3	19	18	37	30	33	63	100
		30	29	59	37	51	88	147
		$\chi^2 abc = 0.0$			$\chi^2 abc = 2.1$			$\chi^2 ab = .9$
High Pre EOL (a_2)	1,2	26	29	55	16	18	34	89
	3	20	15	35	11	8	19	54
		46	44	90	27	26	53	143
		$\chi^2 abc = .5$			$\chi^2 abc = .2$			$\chi^2 ab = 0.0$
								$\chi^2 a = 25.5$***
SMSG								
Low Pre EOL (a_1)	1,2	6	19	25	8	24	32	57
	3	27	21	48	62	26	88	136
		33	40	73	70	50	120	193
		$\chi^2 abc = 5.7$*			$\chi^2 abc = 18.1$***			$\chi^2 ab = .9$
High Pre EOL (a_2)	1,2	15	38	53	13	34	47	100
	3	41	20	61	23	9	32	93
		56	58	114	36	43	70	193
		$\chi^2 abc = 15.7$***			$\chi^2 abc = 13.3$***			$\chi^2 ab = 2.7$
								$\chi^2 a = 18.9$***

* $p < .05$ $\chi^2 abc$ = E - C comparison within sex x pre EOL categories

** $p < .01$ $\chi^2 ab$ = Male - female comparison within pre EOL levels

*** $p < .001$ $\chi^2 a$ = Pre EOL level comparisons

TABLE 18

Ease of Learning index mean r-p scores adjusted for sex and text difficulty judgments.

	Text Difficulty Rank	E low	E high	E ave.	C low	C high	C ave.
Ball State	2,3	40.7	44.6	42.7	42.7	45.8	44.2
	1	52.8	60.1	56.5	54.7	59.0	56.9
	ave.	46.5	52.4	49.5	48.7	52.4	50.5
UICSM	2,3	41.6	45.6	43.6	40.1	43.4	41.8
	1	55.5	59.8	57.6	54.2	60.0	57.1
	ave.	48.6	52.6	50.6	47.1	51.7	49.4
SMSG	2,3	41.9	47.1	44.5	41.6	45.0	43.3
	1	50.2	61.4	55.8	52.3	58.0	55.1
	ave.	46.1	54.2	50.1	46.9	51.5	49.2

Intrinsic Interest index mean r-p scores adjusted for sex and text difficulty judgments.

		E low	E high	E ave.	C low	C high	C ave.
Ball State	2,3	43.1	46.9	45.0	41.8	47.9	44.8
	1	52.4	55.2	53.8	50.2	57.0	53.5
	ave.	47.7	51.0	49.4	46.0	52.4	49.2
UICSM	2,3	45.1	48.6	46.8	43.7	46.6	45.2
	1	50.5	58.9	54.7	51.5	57.0	54.3
	ave.	48.8	53.7	50.8	47.6	51.8	49.7
SMSG	2,3	45.6	53.0	49.3	44.0	52.5	48.2
	1	52.5	60.7	56.6	51.9	57.1	54.5
	ave.	49.1	56.8	52.9	47.9	54.8	51.3

The analysis of variance results are shown in Table 19.

Inspection of the adjusted means in Table 18 compared to those in Table 13 shows that the E-C differences that had been observed on the EOL index were either reduced (Ball State) or tended to favor the E programs (UICSM and SMSG), when sex and text difficulty judgments were taken into account. The analysis across all programs for which a significant treatment difference ($F = 11.5$, $p < .01$) favoring the C class pupils had been observed when text difficulty was not considered, showed a nonsignificant difference ($F = .5$) favoring the E class pupils when the text factor was considered.[10] The analysis within each program comparison condition shows none of the overall E-C differences to be significant whereas for the Ball State program without considering the text judgment a reliable difference ($F = 8.0$, $p < .05$) favoring the C class pupils had been obtained on the EOL index. Also when text difficulty and sex were considered for the SMSG program comparison, a reliable treatment by premeasure interaction was indicated. An additional analysis comparing the SMSG E-C treatment conditions separately for pupils in high and low premeasure categories indicated that among those anticipating relatively less difficulty learning (high pre EOL), E class pupils obtained reliably higher post EOL scores ($F = 3.9$, $p < .05$) than C class pupils. For those in the low pre EOL category the means were in the opposite direction (C>E) but were not reliably different ($F = .56$). A similar difference had not been observed for this program previously. The extent to which scores on the EOL indices are associated with or determined by the degree of textbook difficulty is indicated by the large and highly significant F - ratios obtained for the latter factor.

It is evident that the instructional program differences on the EOL index between pupils instructed with the E and C programs were mainly affected by if not a result of differences in the difficulty of the textbook used for instruction.

Instructional treatment comparisons with adjustments for differences in textbook difficulty responses and pupil sex were also made for the other attitude indices. Analysis procedures similar to those for the EOL scale were used to make E-C comparisons for the Instrinsic Interest, Perceived Utility and Perceived Knowledge r-p index scores. The adjusted means for each of these comparisons are shown respectively in Tables 18 and 20, and the analysis of variance results in Table 19.

For the index of intrinsic interest when text difficulty was considered, no reliable ($p < .05$) E-C treatment differences were indicated for any of the E programs. However for each of the E programs the adjusted means for the E class pupils were larger than those for pupils in the C classes. In contrast, when not adjusted for text difficulty, both the SMSG and Ball State program pupils obtained lower means than their respective C class pupils, the Ball State difference being reliable at the .05 level ($F = 5.0$). It appears that taking the text difficulty differences into account does alter somewhat the E-C difference for the Ball State program. It is also evident that there is a high degree of association between text difficulty judgments and end of year intrinsic interest scores. With respect to sex differences, for all program comparison conditions, boys obtained higher intrinsic interest scores than girls, a difference which was reliable at the .05 level only for the SMSG program.

[10] The results of the analysis over all E program comparison conditions are not shown, only those for the analyses within each E program condition.

TABLE 19

F-ratios from an analysis of variance of each attitude index r-p score considering pupil sex and textbook difficulty judgment.

	Ease of Learning			Intrinsic Interest			Perceived Knowledge			Perceived Utility		
Source of Variance	B.S.	UICSM	SMSG	B.S.	UICSM	SMSG	B.S.	UICSM	SMSG	B.S.	UICSM	SMSG
Treatment	1.1	1.9	1.1	0.0	1.2	2.0	3.3	8.4**	5.2*	.9	5.2*	.1
Text difficulty	198.0***	293.8***	163.2***	82.4***	79.9***	35.4***	75.1***	140.6***	77.9***	32.4***	15.0***	21.0***
Premeasure	24.2***	25.7***	49.5**	25.7***	28.6**	41.2***	51.4***	7.7**	34.5***	14.8***	28.0***	78.0***
Sex	0.0	.3	0.0	1.9	.8	6.6*	.5	0.0	.4	6.0*	8.2**	12.6***
T X t.d.	.4	.6	.1	0.0	.4	.2	.8	.3	.3	1.2	.7	3.6
T X Prem	1.1	.1	3.9*	2.6	.8	.2	.2	.7	.3	1.0	.6	.2
T X S	1.1	.2	1.4	0.0	.3	.1	1.7	2.0	.9	.8	1.0	3.2
t.d. X Prem	1.4	.7	5.3*	0.0	4.0*	.3	.2	2.8	1.4	3.0	.3	[illegible]
t.d. X S	[illegible]	2.6	.3	0.0	2.0	3.1	1.5	0.0	.3	.1	.1	2.7
Prem X S	.5	0.0	1.6	1.7	2.2	0.0	2.6	2.0	.2	.1	.3	.1
T X t.d. X P	.4	.4	1.1	.2	.4	.8	2.6	2.0	.1	.1	0.0	.1
T X t.d. X S	.1	3.6	.4	0.0	1.4	1.6	0.0	1.7	.7	.1	.5	.5
T X P X S	1.0	0.0	1.0	0.0	0.0	.2	.2	3.0	.6	2.3	1.7	.7
t.d. X P X S	1.5	.2	5.0*	.3	0.0	.8	.8	.9	0.0	3.9*	.1	.5
T X t.d. X P X S	4.6*	.2	.1	3.2	3.7	.4	3.0	.3	.2	1.3	.1	1.7

* p <.05, ** p <.01, *** p <.001

TABLE 20

Perceived Utility index mean r-p scores adjusted for sex and text difficulty judgments.

	Text Difficulty Rank	E low	E high	E ave.	C low	C high	C ave.
Ball State	2,3	44.8	47.5	46.1	47.4	48.8	48.1
	1	49.4	56.2	52.8	50.5	54.7	52.6
	ave.	47.1	51.8	49.5	49.0	51.8	50.4
UICSM	2,3	47.5	51.8	49.6	43.2	49.3	46.2
	1	50.2	55.7	52.9	47.9	54.9	51.4
	ave.	48.8	53.8	51.3	45.5	52.1	48.8
SMSG	2,3	44.2	53.3	48.7	45.6	56.0	50.8
	1	51.4	58.6	55.0	49.7	57.1	53.4
	ave.	47.8	55.9	51.9	47.6	56.5	52.1

Perceived Knowledge index mean r-p scores adjusted for sex and text difficulty judgments.

Ball State	2,3	42.6	48.1	45.4	42.4	50.2	46.3
	1	48.3	57.7	53.0	53.0	58.4	55.7
	ave.	45.5	52.9	49.2	47.7	54.3	51.0
UICSM	2,3	44.5	46.2	45.3	41.6	42.2	41.9
	1	55.4	57.6	56.5	50.9	57.5	54.2
	ave.	50.0	51.9	50.9	46.2	49.8	48.0
SMSG	2,3	41.9	50.0	46.0	41.0	47.5	44.2
	1	53.2	58.3	55.8	50.5	55.1	52.8
	ave.	47.6	54.2	50.9	45.7	51.3	48.5

For the Perceived Knowledge index scores, adjustments for differences in textbook difficulty judgments and sex altered the nature of the differences previously observed for each of the E program comparison conditions. When textbook difficulty judgments were considered, the adjusted means for the E class pupils were reliably greater than those of the C class pupils for both the UICSM ($p < .01$) and SMSG ($p < .05$) comparisons. Also when text difficulty was considered for the Ball State comparison, the differences favoring C class pupils were nonsignificant. The latter difference had been significant ($C > E$, $F = 8.2$, $p < .05$) in the earlier analysis shown in Table 15 which did not consider text difficulty while the earlier difference for the SMSG program ($C > E$) had not been reliable. The previous analysis for the UICSM program had not indicated a reliable <u>overall</u> E-C difference but had indicated a significant treatment by premeasure interaction with $E > C$ for those having lower initial perceived knowledge scores and $C > E$ for those with higher initial scores.

These results indicate that when adjustments are made for differences in text difficulty judgments, pupils instructed with the UICSM and SMSG programs tended to perceive their knowledge of math to be greater relative to other subjects than did pupils in the respective comparison classes. Similar comparisons for those instruted with the Ball State program showed no reliable differences in this regard. Here again no reliable sex differences were observed.

There was no real change in the instructional treatment differences obtained for the Perceived Utility scores when pupil sex and text difficulty factors were considered. As observed previously, the UICSM pupils obtained reliably higher scores than their comparison C class pupils while there were no reliable differences for pupils in the Ball State or SMSG programs. This analysis did indicate consistently reliable sex differences across <u>all</u> programs (E and C) with boys tending to see mathematics as having higher utility for future goals than girls. Although textbook difficulty did not apparently differentially affect the Perceived Utility scores of pupils in E and C classes being compared, it is evident from the analysis of variance results that the Perceived Utility scores were related to this judgment about the instructional materials.

The results obtained with adjustments for text difficulty and sex differences were also examined for a measure of general interest in mathematics, the Aiken Interest scale. The adjusted means for each of the E program comparison conditions are shown in Table 21 and the analysis of variance results in Table 22. The means and results of this analysis could be compared to those presented in Tables 2 and 3.

For the Ball State and SMSG programs there is a tendency for the E class means for both sexes to increase relative to the respective C class means when the additional factors were considered. These changes were not large enough however to yield reliable overall treatment differences. For the Ball State comparison a significant treatment by initial interest by text difficulty interaction was indicated. The best characterization of this interaction would seem to be that the degree of text difficulty made a bigger difference for those with higher initial interest in the E classes and for those with lower initial interest in the C classes than for those with lower initial interest in the E and higher initial interest in the C classes.

For the SMSG comparison, a significant sex by treatment by initial interest interaction was indicated which apparently resulted from a higher correspondence between initial and post interest for E class girls and C class boys than for E class boys and C class girls.

TABLE 21

Aiken Interest scale means considering sex and text difficulty for each E program comparison condition.

		E				C			
Ball State	Premeasure Level	low	med	high	ave	low	med	high	ave
Sex	Text Difficulty Rank								
M	3	2.80	3.27	3.72	3.26	2.52	3.02	3.73	3.09
	1,2	2.88	3.65	4.21	3.57	3.12	3.62	4.05	3.60
	ave	2.84	3.45	3.96	3.42	2.82	3.32	3.89	3.34
F	3	2.55	3.18	3.31	3.01	2.39	3.24	3.91	3.18
	1,2	3.06	3.46	4.43	3.65	3.27	3.73	4.16	3.72
	ave	2.81	3.32	3.87	3.33	2.83	3.49	4.04	3.45
All	3	2.68	3.23	3.52	3.14	2.46	3.23	3.82	3.14
	1,2	2.96	3.55	4.32	3.61	3.20	3.68	4.11	3.66
	ave	2.82	3.39	3.92	3.37	2.83	3.40	3.96	3.40
UICSM									
M	3	2.68	2.90	3.13	2.90	2.77	3.18	3.57	3.17
	1,2	2.88	3.67	4.31	3.62	3.29	3.66	4.09	3.68
	ave	2.78	3.29	3.72	3.26	3.03	3.42	3.83	3.43
F	3	2.62	3.42	3.74	3.26	2.32	3.47	3.57	3.12
	1,2	2.90	3.71	4.34	3.65	2.61	3.52	4.15	3.43
	ave	2.76	3.57	4.04	3.46	2.47	3.50	3.86	3.27
All	3	2.65	3.16	3.44	3.08	2.55	3.33	3.57	3.15
	1,2	2.89	3.69	4.33	3.64	2.95	3.59	4.12	3.55
	ave	2.77	3.43	3.88	3.36	2.75	3.46	3.85	3.35
SMSG									
M	3	2.57	2.96	3.94	3.15	2.57	3.07	3.78	3.14
	1,2	3.33	4.00	4.28	3.87	2.74	3.67	4.30	3.57
	ave	2.95	3.48	4.11	3.51	2.66	3.37	4.04	3.36
F	3	2.32	3.23	4.26	3.27	2.43	3.28	3.50	3.07
	1,2	2.70	3.80	4.26	3.59	3.15	3.76	3.98	3.63
	ave	2.51	3.52	4.27	3.43	2.79	3.52	3.74	3.35
All	3	2.45	3.10	4.10	3.21	2.50	3.18	3.64	3.10
	1,2	3.02	3.90	4.28	3.73	2.95	3.72	4.14	3.60
	ave	2.73	3.50	4.19	3.47	2.72	3.45	3.89	3.35

TABLE 22

F-ratios from an analysis of variance of Aiken scale scores considering pupil sex and text book difficulty.

Source of Variation[a]	B.S.	UICSM	SMSG
Treatment	.08	.01	2.57
Text difficulty	37.14***	35.23***	46.56***
Premeasure	62.63***	63.10***	104.13***
Sex	.02	.06	.36
T X t.d.	.09	.84	.02
T X Prem	.02	.07	1.45
T X S	1.36	4.62*	.27
t.d. X Prem	.16	2.33	1.71
t.d. X S	1.19	2.70	.86
Prem X S	.02	3.72*	.95
T X t.d. X P	3.27*	.95	1.43
T X t.d. X S	.87	.17	3.18
T X P X S	.21	.41	3.95*
t.d. X P X S	.73	.45	.57
T X t.d. X P X S	.40	.74	.46

[a]For each source of variation there was 1 degree of freedom.

* $p < .05$, ** $p < .01$, *** $p < .001$

For the UICSM comparison, the E-C differences over all pupils were not altered when sex and text difficulty factors are considered. However, with the adjustment for text difficulty a significant sex by treatment interaction appeared. The Aiken scale mean was greater for girls instructed with the UICSM program than for girls instructed with the conventional program while for boys, those instructed with the conventional program had the higher mean. An analysis of variance for each of the sexes separately indicated however that neither of these differences was very reliable, $F = 2.3$, $.10 < p < .25$ in each instance. In the previous analysis a reliable E-C difference was found only for girls having lower initial interests. The latter difference was not found in the analyses for UICSM girls when text difficulty judgments were considered which suggests that the adjustment for the latter factor worked in this instance in favor of the C class pupils. The analysis for the UICSM program also shows a reliable sex by premeasure interaction which indicates that for girls initial interest was related more highly to post interest than for boys. This fact is also reflected in the higher F-ratio shown for girls compared to boys for the analysis in Table 3.

No clear trends for the E-C differences were revealed on the Aiken scale for the Ball State or SMSG programs when text difficulty judgments and sex were taken into account. However, for the UICSM program, a definite sex by treatment interaction was obtained. Girls in the E classes developed higher interests than those in the C classes while for boys those in the C classes developed the higher interests.

For none of the programs were sex differences indicated on this measure of interest in mathematics. The text difficulty factor did show a consistently reliable effect on Aiken scale scores for pupils in all instructional programs.

When the E-C comparisons were made considering pupil sex and judgments about textbook difficulty, somewhat different instructional treatment difference appeared than had been obtained otherwise for the Aiken Interest scale and the Intrinsic Interest, Ease of Learning and Perceived Knowledge indices. E-C differences on the latter index were especially affected by this adjustment. For each of the measures, the usual effect of the adjustment was to increase the magnitude of the scores obtained by E class pupils relative to those in the respective C classes. This, in effect, indicates that the factors for which the adjustments were being made, sex and text difficulty judgments, were reflecting or contributing to the E-C differences for these measures. Since the adjustment was made for two factors (as well as their interaction) there may be a question as to which factor was the major source of the changes resulting from the adjustment. Examination of the analysis of variance results indicates that text difficulty differences were contributing much more to the variance and therefore to this effect than were the sex differences.

It appears then that when comparisons were made between those who made similar judgments concerning the difficulty of their respective texts, pupils instructed with the E programs exhibited on several indices a more positive or a less negative attitude toward mathematics than those instructed with C programs. The tendency toward development of relatively more positive attitudes appeared to be greater for the UICSM pupils and least for Ball State pupils.

However, before it could be concluded generally that in the absence of text difficulty differences the UICSM and SMSG programs would have had a positive effect on pupils attitudes toward mathematics and the Ball State program would not have contributed to a more negative attitude, other factors must be considered.

It was observed that the Ball State and SMSG pupils had a much stronger tendency to judge their texts as being difficult to understand than did pupils in the C classes. If this were the case, then pupils in the E classes judging their texts as being relatively less difficult were more likely to be pupils that had relatively higher proficiency and/or higher grades in mathematics than the C class pupils making the same judgment. That is the adjustment for text difficulty differences by equally weighting the alternate levels may have resulted in a differential adjustment between E and C classes with respect to mathematics ability or grades received in the mathematics class. Since both of the latter variables, especially the grades received, are relevant to pupil attitudes, partialing out the text difficulty differences could have contributed to more positive attitudes for E class pupils independent of the effects of the instructional materials. This effect would most likely be reflected on the Perceived Knowledge index which correlates most highly with grades. Under these conditions then at each text difficulty judgment level the average grade for E class pupils would be higher than that for C class pupils. If this were observed then pupil grades could account as readily for E-C differences on the Perceived Knowledge and possibly other attitude indices as could the text materials themselves when comparisons were made adjusting for text difficulty differences.

To examine this possibility on the index of perceived knowledge where the greatest variation in results had been obtained, an additional analysis was carried out. E-C comparisons were made considering text difficulty levels controlling for the possible grade and proficiency differences using analysis of covariance. Measures of both of the latter variables and the premeasure of perceived knowledge were included in the analysis as covariates[11]

The results of the covariance analysis indicated no reliable E-C treatment differences for the Ball State and UICSM programs ($F = .86$ and 1.02, respectively). For the SMSG program the treatment difference was reliable, $F_{6,369} = 2.4$, $p < .03$, with the E class pupils having the higher adjusted mean perceived knowledge scores.

Underlying this analysis is the assumption that the covariate regression coefficients were homogeneous across the E-C treatment by text difficulty level categories. A test of this assumption indicated that it clearly held for the UICSM comparison ($F = .57$) but that it was somewhat questionable for the Ball State ($p = .10$) and SMSG ($p = .14$) comparisons which suggests some caution in interpreting the covariance results for the latter two programs.

On the basis of the analysis of covariance it appears than that the E-C differences observed in the UICSM comparison on the Perceived Knowledge index when adjusted for test difficulty differences were primarily the result of proficiency and/or achievement factors. For the Ball State and SMSG programs however, consideration of proficiency and achievement in conjunction with text difficulty judgments levels did not alter the effects observed when text difficulty was considered alone. This suggests for the Ball State program that the E-C differences observed generally (i.e. independent of text difficulty) on the Perceived Knowledge index were associated with or a result of text difficulty judgment differences and for the SMSG program that if the text difficulty judgment differences were adjusted out, the SMSG pupils would obtain higher perceived knowledge scores than those instructed with conventional programs.

[11]A special computer program for the covariance analysis with three covariates was written by Mr. Rodney Rosse.

These results suggest further that the somewhat negative effects observed for the Ball State program and the lack of positive effects for the SMSG program on other attitude indices may have been primarily the result of differences in the difficulty pupils' experienced with the instructional materials.

2. Teacher determined conditions

a. Pupil grades

The grades pupils' receive in a subject are quite likely to have some affect on their attitudes toward the subject matter. Within a given class, grades are no doubt quite highly related to independent measures of the pupil's relative proficiency or ability. It is possible however, independent of actual proficiencies, that teachers might have varied their grading standards between the separate classes they were instructing. Such variation would be more likely when teachers were using different programs of instruction in separate classes which, to the extent that the instructional objectives differ, might require judgments relative to different performance criteria. If this were the case in the present study, it could result in a difference in the distribution of grades assigned respectively to the E and C class pupils. There is a question then as to whether there was any systematic difference in the grades assigned to pupils in the E and C classes when any previous mathematics class performance differences were held constant.

Information concerning the average or overall grade received in mathematics during the previous (8th grade) and concurrent year was obtained for pupils in a majority of classes participating in the study. To determine if there were any grade differences between the E and C class pupils, comparisons were made within each E program comparison condition using the three factor analysis of variance. The analysis was restricted to class pairs for which the necessary data was available and this consequently reduced the number that could be included in each program comparison. Table 23 shows the adjusted grade means for pupils instructed with the alternate programs and Table 24 shows the results of the analyses of variance for this measure. The analyses provided no indication that there were any overall differences in the grades received by pupils in the E and C classes for each E program comparison condition. It does not appear then that any instructional treatment differences observed with respect to pupil attitudes could be attributable to general or systematic differences in grades received by E and C class pupils respectively.

A moderately reliable teacher by treatment interaction was observed for both the Ball State and UICSM program comparisons. This result indicates some variation in the magnitude and/or direction of the E-C grade differences between teachers. Whether the teacher by treatment grade differences correspond to between teacher treatment differences on other measures will be considered in a subsequent analysis.

b. Teacher attitudes and judgments concerning the E programs

Another instructional factor or condition which could have differentially affected the attitudes of pupils in a pair of E and C classes was the judgment or attitude of the teacher concerning the materials being used. A characterization of the materials by the teacher to the pupils, either in general evaluative terms or in terms of their anticipated instructional outcomes or purposes could affect the pupils reaction either to the materials or to the subject matter itself. Such a characterization

TABLE 23

Adjusted grade means for E and C class pupils in each E program comparison codition.

Experimental program	Number of class pairs	Premeasure level	E low	E high	E ave.	C low	C high	C ave.
Ball State	8		3.14	5.36	4.25	3.28	5.15	4.22
UICSM	5		3.02	5.42	4.22	3.75	5.74	4.74
SMSG	8		3.96	5.48	4.72	3.45	5.48	4.46

Values assigned letter grades:

A = 1
A-, B+ = 2
B = 3
B-, C+ = 4
C = 5
C-, D+ = 6
D = 7
D- = 8
F = 9

TABLE 24

F-ratios from the analysis of variance for pupil grades in each program comparison condition.

Source of Variance	d.f.	B.S.	UICSM	SMSG
Treatment	1	0.0	1.4	2.3
Premeasure	1	119.1***	21.0*	52.7***
Teacher	$(t - 1)^a$	6.4***	17.6***	6.2***
Treat X Premeasure	1	.9	.7	2.2
Treat X Teacher	(t - 1)	2.5*	3.0*	1.2
Premeasure X Teacher	(t - 1)	.6	3.7**	2.0*
Treat X Premeasure X Teacher	(t - 1)	.9	.2	.5

[a] t = number of class pairs for each program condition indicated in Table 23.

* p <.05, ** p <.01, *** p <.001

would be especially likely in the E classes because the materials were obviously different. As part of the more extensive project concerned with the effects of the E programs on pupil achievement in mathematics, a questionnaire had been prepared to elicit participant teacher judgments and reaction to the experimental instructional materials they were using. From the questionnaire items several scales were derived. The questionnaire provided indices of (1) the teacher judgment of the relative extent to which the E and C programs achieved certain instructional objectives. (2) The teachers preference for E relative to C materials and (3) the teachers judgment of the pupils reaction (in general evaluative terms) to the materials and the subject matter.

To examine this question, a single index representing the E-C class difference (adjusted for premeasure differences) on each of the attitude scales for each individual teacher was obtained. The degree of association between the teachers adjusted E-C class difference score for each attitude index and teachers score on each of the teacher judgment indices was determined. Classifying teachers above or below the median on each pair of indices, the degree of association was determined by an exact probability test. The results indicated that none of the teacher judgment indices were reliably related to the instructional treatment (E-C) differences obtained for individual teachers for any of the attitude indices. That is, there was no correspondence between the individual teachers judgments concerning the E programs and the E-C class differences on any of the pupil attitude indices. It appears then that there is no evidence that the teacher attitudes or judgments as measured by the teacher questionnaire indices were related to or affected the differences in attitude observed between the E and C classes.

C. Instructional treatment effects considering moderator variables

Another question separate from that concerning the general effects observed for the E programs, is the question of whether the effects on attitudes and interests are the same for all pupils or the same under all conditions of instruction. This is a question of whether certain pupil characteristics or instructional conditions function as moderator variables, in the sense that they interact with the instructional program variations, to alter or modify their effects.

On the basis of more general considerations, somewhat different attitude and interest effects might be expected for such pupil characteristics as sex or level of mathematics ability or for classes of teachers varying in the amount of experience they had had using their respective E programs. Analyses were carried out considering each of these factors as a possible source of differential attitudinal effects for the separate instructional programs.

1. Pupil characteristics

In the previous analyses sex differences in attitudes toward mathematics as they developed over the year were observed in several instances. On the Perceived Utility index a consistently higher score was obtained by boys which was however independent of the various instructional program differences and no doubt reflected the effects of factors other than the instructional materials. Although the instructional treatment effects tended to be in the same direction for both sexes, larger and more reliable differences were observed more frequently for girls in the analyses using the a-v scores on the Intrinsic Interest, Utility and Ease of Learning indices.

Another indication of pupil sex as a moderator variable was observed on the Aiken scale for the UICSM comparison. This appeared in the initial analysis for the Aiken scale shown in Table 4 for which a reliable treatment by premeasure effect was observed for girls but not for boys. It also appeared when E-C comparisons were made on the Aiken scale adjusting for text difficulty differences. The latter analysis indicated that girls instructed with the UICSM program developed more positive interests than those in the comparison C classes while the boys showed an E-C difference of similar magnitude in the opposite direction.

With respect to mathematics ability or proficiency as a possible moderating variable, the question is whether the E programs affect the attitudes of pupils having relatively higher and lower proficiency in any differential way. This question derives from logical as well as empirical considerations. In a separate questionnaire, distributed to obtain their reactions to the experimental programs they were teaching and judgments of their pupil's reaction to these materials, teachers were asked to indicate, for high, average, and low ability pupils separately, whether those in the E or C classes responded more favorably to their respective materials. A high proportion of the responding teachers indicated that among higher ability pupils, the response was more favorable for those in the experimental class, while among low ability pupils the response was more favorable for those in the C classes. (See Ryan and Rising (9)).

The judgment probably represents a belief that the somewhat greater emphasis on the conceptual aspects of mathematics in the E programs would have relatively less appeal to lower ability pupils, while the somewhat more rote computational and rule learning character of the conventional programs would have relatively less appeal to the higher ability pupils. If true, then an interaction between ability and instructional program should be indicated on a measure of interest such that among lower ability pupils those in the E classes would have less interest at the end of the year than those in the C classes, while among higher ability pupils, those in the E classes would have the greater interest.

To examine this question in general, comparisons were made for each of the several attitude and interest indices; the Aiken scale - one of the measures of general interest, and the Intrinsic Interest, Perceived Utility and Perceived Knowledge indices. Since as discussed above, some evidence has appeared indicating that the treatment effects on pupil interest might be modified by sex differences, pupil sex was also included in the analysis to examine this possibility further and to determine whether there was a sex by ability interaction.

The pupils score on the mathematics section of the STEP obtained at the beginning of the year served as a measure of mathematics ability or proficiency. The comparisons were made using analysis of variance within each E program condition. Four factors were considered; instructional treatment (E or C), pupil sex, and initial (pre) levels of proficiency and of interest, the latter being the premeasure of the dependent variable. For each of the latter two measures, two levels determined by the median of the distribution of scores of all pupils on each were used.

In this analysis, if the treatment effects did vary generally with the pupils' initial level of proficiency, with sex or with both factors, this would be indicated by significant treatment by proficiency, treatment by sex, or treatment by proficiency by sex interactions, respectively. Table 25 shows the adjusted means on each of the attitude indices for pupils having higher and lower levels of proficiency within sex,

TABLE 25

Adjusted interest index means for E and C class pupils by sex and level of proficiency in math for each program comparison condition.

Ball State	Math Proficiency Level	Aiken E	Aiken C	Intrinsic Interest E	Intrinsic Interest C	Perceived Utility E	Perceived Utility C
Sex	low	3.13	3.23	46.4	49.4	47.4	52.1
M	high	3.58	3.59	49.2	51.7	51.3	53.5
	ave.	3.35	3.41	47.8	50.6	49.3	52.8
	low	3.02	3.37	46.1	48.0	48.1	48.3
F	high	3.53	3.66	49.0	51.8	47.0	51.0
	ave.	3.27	3.51	47.6	49.9	47.5	49.7
	low	3.08	3.30	46.2	48.7	47.7	50.2
All	high	3.55	3.62	49.1	51.8	49.1	52.3
	ave.	3.31	3.46	47.7	50.2	48.4	51.2
UICSM							
	low	3.10	3.23	48.7	49.4	51.7	49.1
M	high	3.45	3.64	52.4	51.5	53.2	53.2
	ave.	3.27	3.43	50.5	50.4	52.4	51.1
	low	3.36	3.16	48.2	49.0	49.4	45.4
F	high	3.51	3.52	52.1	50.7	50.8	50.9
	ave.	3.44	3.34	50.1	49.9	50.1	48.1
	low	3.23	3.19	48.4	49.2	50.5	47.2
All	high	3.48	3.58	52.2	51.1	52.0	52.1
	ave.	3.35	3.39	50.3	50.1	51.3	49.6
SMSG							
	low	3.10	3.43	48.4	51.0	50.2	52.5
M	high	3.52	3.40	54.3	54.5	52.5	53.4
	ave.	3.30	3.41	51.4	52.7	51.3	52.9
	low	3.11	3.31	49.2	51.2	47.3	52.6
F	high	3.23	3.44	51.9	48.6	50.1	50.3
	ave.	3.17	3.37	50.6	49.9	48.7	51.4
	low	3.12	3.37	48.8	51.5	48.7	52.5
All	high	3.37	3.42	53.1	51.6	51.3	51.8
	ave.	3.24	3.39	50.9	51.3	50.0	52.2

instructional treatment, and E comparison condition categories. Tables 26 and 27 show the results of the analyses of variance for each of the indices.

As can be seen, for none of the interest measures were the specific lower order interactions indicating a general moderating effect for pupil proficiency and/or sex significant at the .05 level or less. There was a tendency for the differences to be more in favor of the C programs among those with lower than higher proficiency for the Ball State pupils and SMSG males on the Aiken scale. A similar tendency was noted for the SMSG comparison on the Intrinsic Interest and Perceived Utility indices. For the UICSM program on the Perceived Utility index the trend was in the opposite direction, i.e., the E class mean being relatively higher for the lower proficiency pupils. None of these differences, however, were large enough to yield reliable effects.

A variation between the E-C differences for the sexes for the UICSM program comparison on the Aiken scale which was noted above also appeared in this analysis but was not significant. Apparently the adjustment for initial proficiency differences did not affect the sex by treatment interaction in quite the same way as did the adjustment for text difficulty differences.

For the Aiken scale, two higher order interactions were indicated; a treatment by proficiency by premeasure interaction for the UICSM program comparison and an interaction involving all four factors for the Ball State program. Neither of these interactions appeared to reflect an easily interpretable pattern of effects[12] and since the analysis was concerned primarily with more general, i.e. lower order, interaction effects no further comparisons were carried out.

In general, on the basis of this analysis, it does not appear that the effects of the experimental programs vary between pupils having different levels of proficiency in mathematics at the beginning of the year, i.e., mathematics proficiency does not moderate the instructional program differences.

This analysis also did not provide any more conclusive evidence that there were any differential treatment effects associated with sex differences. The tendency for a sex by treatment interaction in the UICSM program comparison condition on the Aiken scale was consistent with a separate analysis but did not appear in this instance to be very reliable. This effect as observed in the previous analysis may, however, have been a result of factors discussed below.

The analyses considering initial pupil ability and sex, however, revealed other differences which have more general implications concerning the instructional treatment effects. It can be seen in Table 26 for the SMSG program comparison on the Aiken scale that a reliable treatment difference was indicated which was the result of higher mean scores obtained by pupils instructed with the conventional rather than the SMSG program. This difference had not been found to be as large nor as

[12]For the UICSM program, in terms of E-C differences the interaction reflected a difference in favor of the E class pupils for those low on premeasures of both ability and interest or high on both premeasures while the difference favored C class pupils in the remaining two cross classification categories.

TABLE 26

Results of analysis of variance of Aiken scale and Intrinsic Interest index (r-p) scores within each experimental program condition considering pupil sex and proficiency in mathematics.

		Aiken						Intrinsic Interest					
		Ball State		UICSM		SMSG		Ball State		UICSM		SMSG	
Source of Variation	d.f.	Mean square	F	Mean square	F	Mean square	F	Mean square	F	Mean square	F	Mean square	F
Treatment	1	.0877	3.4	.0042	.2	.0979	4.6*	26.4	6.8**	.1	.0	.5	.1
Sex	1	.0004	.0	.0050	.2	.0306	1.5	.8	.2	1.0	.2	13.4	3.3
Math Proficiency	1	.6385	24.5***	.4029	14.9***	.0947	4.5*	34.9	9.0**	31.9	7.3**	22.3	5.4*
Premeasure	1	2.4120	92.4***	3.1732	117.4***	4.0574	192.0***	148.9	38.3***	129.6	29.7***	272.9	66.5***
T X S	1	.0333	1.3	.0652	2.4	.0087	.4	.2	.1	.0	.0	4.2	1.0
T X MP	1	.0240	.9	.0175	.7	.0479	2.3	.0	.0	3.9	.9	14.1	3.5
T X P	1	.0022	.1	.0040	.2	.0014	.1	5.1	1.3	.8	.2	.1	.0
T X S X MP	1	.0049	.2	.0062	.2	.0515	.4	.6	.2	.1	.0	2.0	.5
T X S X P	1	.0190	.7	.0000	.0	.0187	.9	.5	.1	.2	.1	1.3	.3
T X MP X P	1	.0029	.1	.1263	4.7*	.0277	1.3	1.9	.5	4.9	1.1	1.6	.4
T X S X MP X P	1	.1048	4.0*	.0012	.0	.0011	.1	11.9	3.1	7.4	1.7	1.2	.3
Residual	4	.0442	1.7	.0261	1.0	.0257	.3	4.8	1.9	5.6	1.3	7.3	1.8
Adjusted error	d.f.	341		267		392		342		317		340	
	M.S.	.0260		.0270		.0211		3.9		4.4		4.1	

TABLE 27

Results of analysis of variance of Perceived Utility index (r-p) scores within each experimental program condition considering pupil sex and proficiency in mathematics.

Source of Variation	d.f.	Ball State Mean square	Ball State F	UICSM Mean square	UICSM F	SMSG Mean square	SMSG F
Treatment	1	31.6	8.4**	10.8	2.3	18.5	6.3*
Sex	1	24.4	6.5*	28.4	6.0*	16.9	5.7*
Math Proficiency	1	11.7	3.1	39.3	8.2**	3.4	1.2
Premeasure	1	60.7	16.1***	159.0	33.3***	305.9	103.3***
T X S	1	1.9	.5	.5	.1	1.4	.5
T X MP	1	.4	.1	11.0	2.3	10.5	3.5
T X P	1	.3	.1	6.9	1.5	.3	.1
T X S X MP	1	10.5	2.8	.7	.2	3.4	1.1
T X S X P	1	5.1	1.4	9.8	2.1	3.3	1.1
T X MP X P	1	4.3	1.2	.1	.0	1.6	.6
T X S X MP X P	1	1.5	.4	3.6	.8	.0	.0
Residual	4	4.1	1.1	1.9	.4	2.5	.8
Adjusted error	d.f.	396		321		399	
	M.S.	3.8		4.8		3.0	

reliable in the previous analysis (shown in Tables 2 and 4) which had not considered (i.e. adjusted for possible differences associated with) pupil ability or sex, but had considered teacher differences. Moreover, a change of some degree in the magnitude and reliability of the instructional treatment differences for the E programs was also indicated for the index of perceived utility using r-p scores when pupil sex and initial proficiency were considered. For the latter analysis, reliable E-C differences were indicated in the Ball State and SMSG comparisons with the C class pupils having the higher mean scores in both instances. Although in the previous analysis (shown in Table 15) these differences were in the same direction, they were not statistically reliable. For the UICSM program in the analysis (shown in Table 27) the E-C difference was smaller and less reliable (actually non-significant by the criterion being used) than had appeared in the previous analysis.

The difference in outcome between the two analyses could be due to either or both of two conditions:

1) A variation in the E-C differences for individual teachers for whom there was also a concommitant variation in class size. The earlier analysis adjusted for class size differences (by considering "teacher" as a dimension) while that involving pupil ability and sex did not include this adjustment. However, if between teacher variations in E-C differences were of some magnitude, this should have been reflected by a reliable teacher by treatment interaction.

2) A difference between E and C class pupils with respect to the distribution of one or both of the factors (proficiency and/or sex) for which there was some degree of correspondence with the dependent variable that had not been adjusted out by the premeasure control variable. The procedures followed for the analysis were directed toward minimizing E-C differences with respect to the premeasure for a given index. These procedures should also have reduced or minimized E-C differences with respect to any other beginning-of-year pupil characteristics which happened to be correlated with the dependent variable assuming that they would be at least as highly correlated with the premeasure of the dependent variable. It may have been that the analysis procedures did not provide a sufficient control or adjustment for the effects of certain pupil characteristics such as sex or initial proficiency on at least some of the indices and/or these characteristics were more highly related to the post than the premeasures of the variable.

It is on the Perceived Utility index that the alternate analyses differ most in the interpretations they would permit with respect to the reliability (but not direction) of the differences for the separate E program. On this index for the UICSM comparison both proficiency and sex appear to relate to or affect the dependent variable as indicated by the main effects for each, while no teacher by treatment interaction was indicated. This leaves the sex and ability differences as the most plausible explanation for differences obtained by the separate analyses for the UICSM program comparison.

The latter possiblity is supported further by a closer examination of the data which revealed a slightly higher overall proportion of both males and higher proficiency pupils (members of both categories tending to have higher scores) in the UICSM classes.

For the SMSG and Ball State comparisons, reliable teacher by treatment interactions as well as sex differences were obtained. For each of these comparisons, the distribution with respect to proficiency was quite similar for E and C groups but for

the Ball State comparison there was a higher proportion of males in the C classes. Consequently for these programs either of the above conditions may have contributed to the variation in results.

The occurrence of an overall difference in the sex and proficiency distributions however does not unequivocally indicate that these factors are accounting for the variation in treatment effects indicated by the analyses since the analysis design could have controlled for some variation in these factors. A more extensive analysis providing for direct control on the possible effects of each of these factors would be required to determine the nature of the E-C differences in more precise terms.

For the most part the results observed with respect to the treatment differences when pupil sex and proficiency are taken into account affect primarily the conclusions that can be made for the UICSM program on the Perceived Utility index.[13] It appears that the reliably higher score indicated for UICSM instructed pupils in the previous analysis may have been due in part to factors other than the instructional program. On the other hand, for an analysis on Perceived Utility scores across all E program conditions, a reliable treatment by program interaction ($F_{2,1116} = 6.1$, $p < .01$) was again obtained when pupil sex and proficiency were considered. The latter interaction reflected the variation in the direction of the E-C differences between the UICSM and the other E program comparison conditions which had been observed in the previous analysis shown in Table 14.

2. Teacher experience with the experimental programs

It would be reasonable to expect that as teachers had additional years experience with a new program of instruction, such as the E programs, they would be in a better position to impliment the instructional objectives specific to that program and probably reduce somewhat any special difficulties or additional effort required in connection with its use. To the extent that such factors affected pupil attitudes either directly or indirectly, they would contribute in general to greater variation in E-C differences among teachers having differing amounts of experience and specifically to greater differences in favor of the E program for classes of teachers with more experience.

On the other hand, however, over time there might be a tendency, due in part to increased familiarity, for the teacher to introduce some of the more positive characteristics of the E programs in his conventional classes. If this were the case smaller E-C differences in pupil outcomes affected by these factors would be expected. A question exists therefore, as to whether E-C differences for the measures obtained did vary between teachers having more or less experience with the respective E programs they were using, and if so, whether there were reliable E-C differences for classes of teachers at one experience level which had not been indicated generally.

In the previous analyses for each of the attitude indices, individual teachers were treated as levels on a separate dimension. Consequently if any teacher connected charateristics contributed to a reliable variation in the treatment differences this

[13] A similar analysis was also carried out for the Perceived Knowledge index with the results being very similar to those obtained for the earlier analysis shown in Table 15.

effect would necessarily be reflected in the analyses by a significant teacher by treatment or teacher by treatment by premeasure interaction. Conversely, nonsignificant teacher by treatment interactions would indicate that there were no real variations across teachers with respect to E-C differences which could be attributed in teacher characteristics.

Examining the results of the analysis carried out for the separate attitude measures in which the teacher effect was treated as a separate dimension, only for the following measures and E program comparison conditions were significant teacher by treatment interactions indicated: Perceived Knowledge, s-r score, Ball State; Perceived Utility, r-p score, Ball State and SMSG, and Ease of Learning, r-p score, SMSG.[14]

Where the previous analyses indicated a significant teacher by treatment interaction, additional analyses were carried out to determine if the between teacher variation in E-C differences resulted from or was associated with differences in teacher experience with the E programs.

For each of the above measures and E program conditions, teachers were classified according to their relative level of experience and the reliability of the E-C treatment difference within each experience level was determined using analysis of variance within each experience level. The adjusted means being compared and the pertinent results of the analysis are shown in Table 28.

For the Ball State and UICSM programs on the Perceived Utility index (r-p scores) and the Ball State program on the Perceived Knowledge index (a-v scores), none of the within experience level treatment differences reached the .05 level of significance. It is evident though that in each of these instances the classes of the most experienced teachers tended to exhibit the smallest E-C differences. The within experience level comparison for the SMSG program on the EOL index (r-p scores) did however show a highly reliable treatment difference for classes of teachers having the most experience with this program. The latter result indicates that the tendency observed more generally for E class pupils to experience greater learning difficulty occurred for SMSG instructed pupils in classes of teachers having the most experience with this program. On logical grounds it would seem that a difference of this type would be more likely among classes of teachers having less rather than more experience with a specific program.

D. Properties and relations among attitude & interest indices and other measures

As discussed above there are some general questions of a methodological nature that should be considered concerning the various attitude and interest measures used in the study which are relevant to the effects observed.

There is a question of reliability or how well or consistently the indices provided measures of the characteristics they were intended to assess.

There is also the question of validity of the indices used. In the context of the objectives of this study this question concerns the degree to which the separate

[14]The teacher by treatment effect was also obtained on the Aiken scale for girls alone. However this effect was not examined further since the teacher experience factor would be presumed to be a factor affecting both sexes.

TABLE 28

Results of analysis of variance within levels of teacher experience with E program.

Index	E program	years exper.	No. of tchrs.	Treatment adjust. means E	C	differ. E - C	Treatment M.S.	d.f.	Error M.S.	d.f.	F ratio
Perceived Utility	Ball State	1,2	3	47.1	51.0	-3.9	45.5	1	24.3	2	1.9
		3	5	48.5	49.8	-1.3	8.6	1	20.5	4	.4
Perceived Utility	SMSG	1	5	49.2	54.0	-4.8	112.4	1	30.5	4	3.7
		2,3	3	49.6	51.2	-1.6	7.8	1	6.6	141	1.2
Ease of Learning	SMSG	1	5	47.1	49.3	-2.2	24.5	1	71.6	4	.3
		2,3	3	46.4	51.1	-4.7	66.4	1	6.6	141	10.0***
Perceived Knowledge	low premeasure	1,2	5	8.1	9.2	-1.1	3.33	1	.96	4	3.5
Ball State		3	3	8.1	7.9	.2	.09	1	1.92	2	.1
	high premeasure	1,2	10.6		10.4	.2					
		3	10.3		10.6	-.3					

A preliminary test indicated no reliable variation in E - C differences between teacher experience levels, therefore no within level tests were run.

attitude and interest indices are measuring relatively independent of separate attitudinal reactions or dimensions and whether on the whole the affective factors or reactions being measured are independent of performance or achievement characteristics and more general academic attitudes.

1. Reliability

Data were gathered and analyses carried out to obtain estimates of both the internal consistency and the stability of the scores for the various indices or measures that were used.

To obtain an estimate of the reliability of the various attitude indices over time, the Fall version of the questionnaire containing all items in the indices was administered twice within a three month interval to a sample of 200 ninth grade algebra pupils in three schools not participating in the main study. Product-moment correlations were computed for both absolute-value and rank-position scores on the indices derived from the questionnaire responses. The test-retest reliability coefficients for the "retest" sample are shown in Table 30. (A separate indication of the stability of the indices over the school year was obtained for the data sample from the correlations between scores on the same indices obtained from the Fall and Spring questionnaires. These are shown in Table 31).

Internal consistency coefficients for the various indices were determined using the Hoyt reliability formulas found in Cronbach, Rajaratnam and Gleser (5).[15]

Internal consistency reliability coefficients were computed separately for the indices from the first and second administration of the questionnaire (Fall version) to the retest sample shown in Table 30 and for the indices derived from the Fall and Spring versions of the questionnaire for the actual data sample shown in Table 29. These coefficients were computed for both absolute value and rank position (3 subjects; English, science and mathematics) scores. The reliability coefficients indicate that each of the separate indices was sufficiently internally consistent to provide an adequate measure of a pupil characteristic for purposes of group comparison.

2. Relations between indices

Tables 31 and 32 show the correlations obtained between the separate attitude and interest indices and measures of achievement for the absolute value and rank position scores respectively.

Considering the correlations at the beginning of the year for the absolute-value scores, the pattern of empirical relations among these variables appears to fit a logical classification which would distinguish three separate categories of variables: (1) External indices of proficiency in mathematics - grade and achievement test scores. (2) Attitudes toward the subject matter per se- indices of intrinsic

15 This procedure utilizes analysis of variance considering the within pupil item response variance and the between pupil score variance.

The Hoyt formula is:

$$r_{11} = \frac{\text{MS between pupils} - \text{MS within pupils}}{\text{MS between pupils}}$$

TABLE 29

Coefficients of internal consistency obtained for the attitude and interest measures from the data sample (N = 1100).

Measure	Fall		Spring	
	a-v	r-p	a-v	r-p
Aiken scale	a	-	96	-
Intrinsic Interest	78	80	81	86
Perceived Utility	77	75	83	75
Perceived Knowledge	72	76	78	86
Ease of Learning	73	71	89	87

[a]Computed only for spring administration

TABLE 30

Reliability coefficients obtained for the attitude and interest indices from the "retest" sample.

Measure	Test-retest		Internal Consistency			
			First Admin.		Second Admin.	
	a-v	r-p	a-v	r-p	a-v	r-p
Aiken scale	81	-	a	-	a	-
Dutton scale	64	-	b	-	b	-
Intrinsic Interest	70	60	80	83	82	86
Perceived Utility	54	50	72	76	84	81
Perceived Knowledge	61	65	79	91	82	90
Ease of Learning	68	61	76	71	75	86

[a]An internal consistency coefficient for the Aiken scale was determined for the data sample only since this scale had been originally developed independent of the present data sample.

[b]Because of the nature of the response required for the separate items on this Likert type scale, it was not appropriate to compute the coefficient used for the other scales. No alternative procedure for computing the internal consistency for this scale was determined.

TABLE 31

Correlations between absolute-value scores obtained on attitude and interest indices and measures of achievement.*

Measure	1	2	3	4	5	6	7	8	9
1 Intrinsic Interest	55	54	70	67	54	59	36	36	22
2 Perceived Utility	42	42	51	49	43	43	30	28	29
3 Aiken scale	67	39	69	85	68	70	48	48	34
4 Dutton scale	60	38	82	65	61	63	44	42	32
5 Perceived Knowledge	41	33	57	52	59	72	57	68	45
6 Ease of Learning	47	29	58	51	62	57	56	58	39
7 Expected Grades	41	28	51	46	63	56	45	68	39
8 Actual Grades	17	14	28	25	45	31	50	65	49
9 STEP-Math	07	15	23	22	38	27	38	43	61

TABLE 32

Correlations between rank-position scores obtained on attitude and interest indices and measures of achievement.*

Measure	1	2	3	4	5	6	7	8	9
1 Intrinsic Interest	47	56	59	57	69	60	53	32	19
2 Perceived Utility	50	39	36	36	43	35	33	10	13
3 Aiken scale	47	24	69	85	56	52	47	48	34
4 Dutton scale	45	25	82	65	56	50	44	42	32
5 Perceived Knowledge	67	38	45	43	44	78	66	43	19
6 Ease of Learning	47	34	44	42	67	45	67	39	15
7 Expected Grades	52	35	44	41	67	63	37	38	17
8 Actual Grades	12	02	31	28	15	15	20	65	49
9 STEP-Math	08	02	23	22	13	14	15	43	61

*Fall (pre) scores shown below diagonal

Spring (post) scores shown above diagonal

Fall-spring correlations for same scales shown on the diagonal

interest, perceived utility and general attitudes and interests in mathematics. (3) Pupil judgments or perceptions of their own proficiency or ability to successfully achieve the goals implicit in the subject matter - indices of ease-of-learning, perceived knowledge and expected grades. In the correlation tables the variables have been grouped according to these categories.

This grouping is supported by the fact that the interrelations among the attitude and interest indices tend to be consistently higher than the relations between these indices and the separately obtained objective indices or proficiency - pupil grades and achievement test scores. For the STEP mathematics test, which on logical grounds would seem to be the best index among these measures of the pupils' mathematics proficiency or ability, the lowest relations were obtained. It is evident that the attitude and interest indices on the whole were not reflecting to any great extent factors in common with the pupils' measured proficiency or classroom performance in mathematics. That is, the relations obtained indicate that the indices of pupil attitudes toward mathematics subject matter were assessing qualities that were relatively independent of objective measures of ability in mathematics. Also, the variables reflecting the pupils' subjective impression of his ability had higher relations with both subject matter interest and actual ability than did the latter two sets of variables with each other. However, the former set of variables appeared sufficiently independent to be treated as representing conceptually separate qualities or characteristics of the pupils. At the end of the year, the relations obtained for the absolute-value scores generally tend to be higher and the subgroupings less clear. The correlations of the other measures with actual grades and with the EOL index show the greatest increase as does the intercorrelation between these two variables.

This appears to suggest that over the year the factors associated with pupil grades and learning difficulty come to have a strong affect on the other variables. Whether the distinctions between the separate categories of variables as well as between measures of variables within the categories are empirically useful and valid will be determined by the increased prediction and differentiation such distinctions can provide which will be indicated in subsequent analyses to be carried out with this data concerning questions of a more general nature. That the distinction, for example, between measured and perceived ability or proficiency in school subjects in general can yield differential and meaningful relations has been shown in recent research by Brookover (4). Also some indication is provided by the results presented above which reveal a variation in the treatment effects observed on the separate indices for the different E program comparisons. These results do not however provide unequivocal evidence that all of the measures included in this study were sufficiently independent to have provided functionally unique information concerning the instructional outcomes.

The relations among the indices having rank-position scores when these scores are considered (Table 32) tended to be somewhat higher than obtained for the a-v scores. This indicates that the r-p scores were reflecting more common factors than were the a-v scores. It also appears that the r-p scores had lower correlations than a-v scores with measures for which only a-v scores were used (i.e. Aiken and Dutton scales, grades and test scores) indicating that the factor or factors partialed out of the a-v scores (i.e. not reflected in the r-p scores) were contributing to higher correlations among some of the a-v scores, i.e. that there was a common a-v score factor. Additional analyses are necessary to obtain a better picture of the factors contributing to the relations obtained between these variables for the

separate scores that were used and between the r-p and a-v scores.

It might be noted that the pre-post correlations for each of the measures were lower for the r-p than the a-v scores which indicates that the r-p scores had a greater tendency to change. Since r-p and a-v scores show about the same level of internal consistency, the r-p scores exhibit somewhat more than the a-v scores the qualities Bereiter (3) has suggested as desireable for assessing change, e.g. high internal consistency, low correlation over time.

IV. Discussion

The question of main concern for this study was whether the experimental programs contributed to the development of differential pupil attitudes toward mathematics as compared to conventional programs of instruction.

Considering the most general measures of interest in and attitude toward mathematics, as provided by the Aiken and Dutton scales, there appeared to be no indication of an overall differential effect for the E programs. Only for the UICSM program was any difference observed for these measures and this resulted from a difference for girls having initially lower interests. Among the latter, those instructed with the UICSM program did exhibit higher interests than those instructed by the same teachers with conventional programs.

On the measure developed to assess specific attitude and interest dimensions, somewhat more general and definite differences among the instructional programs were observed in the main analysis.

Considering the most direct indices of attitudes, for the comparisons involving the Ball State program there was a fairly consistent tendency for pupils so instructed to develop less positive attitudes than pupils instructed with conventional programs. This was most clearly indicated for the a-v scores on the Perceived Utility index and for the r-p scores on the Intrinsic Interest and Perceived Knowledge indices.

At the same time for the UICSM instructed pupils there were indications of the development of somewhat more positive attitudes than for those in the conventional program classes taught by the same teachers. This was most clearly indicated on the Perceived Utility index for both a-v and r-p scores. Similar differences for girls alone were found on the Intrinsic Interest index.

In general considering the results from the main analyses for all of the attitude indices there were some indications that the E programs effected the development of differential attitudes. These effects tended toward less positive attitudes for the pupils instructed with Ball State program and toward more positive attitudes for those instructed with the UICSM program with no very consistent differences observed for the SMSG program. In addition there were indications that pupils instructed with each of the E programs experienced more difficulty learning the subject matter than pupils instructed with conventional programs. For the most part, however, none of the differences was very large, accounting for a relatively small proportion of overall score variance in each instance.

Subsequent analyses were carried out to determine the extent that attitude differences between the experimental and conventional programs resulted from or were influenced by concomitant instructional factors or conditions or certain pupil characteristics. In addition to indicating the probable basis for some program differences, these analysis suggested that some qualification of the UICSM differences was necessary.

Examination of several instructional factors to determine their contribution to the attitudinal outcomes revealed large instructional program differences in pupil judgments of their instructional materials. Both the Ball State and SMSG instructed pupils reported relatively greater difficulty understanding their texts much more frequently than did conventional class pupils. The UICSM pupils, however, did not differ in this respect from the C class pupils with whom they were compared.

The results also indicated that "textbook difficulty" was a factor affecting end-of-year scores on the attitude indices and further that this factor contributed to a more negative attitude than would otherwise have been observed for the Ball State and the SMSG instructed pupils.

Separate analyses considered grading differences and indices reflecting teacher evaluations of the experimental program they were teaching as factors contributing to E-C differences. No evidence was obtained to indicate that either of these factors differentially affected the attitudes of pupils in any of the E or C groups being compared.

The effect of the teacher's experience with the E programs was also examined. The only difference observed was on the ease of learning index for the SMSG teachers where greater learning difficulty among E class pupils was observed for teachers who had the greatest amount of experience with the E program. This result is somewhat difficult to explain and probably should be supported by further evidence before it is considered to be a general effect associated with the SMSG program.

Comparisons were also made to determine whether instructional program differences varied with certain relevant pupil characteristics such as sex or proficiency in mathematics. From the analyses carried out, there was no evidence that the E programs had any differential effects on the attitudes of pupils of higher or lower levels of proficiency in mathematics using as a measure of proficiency test scores obtained at the beginning of the year.

Although there appeared to be indications that the instructional program effects varied between sexes for certain attitude indices when a-v scores were used, similar indications were not obtained when sex differences were assessed directly for r-p scores on the same index. It is possible that there was a sex difference in response style or other factors specific to the a-v scores.

Overall the largest instructional program effects were observed for the pupils' judgments concerning the difficulty of the instructional materials. However, in spite of the rather large differences in this regard for pupils instructed with the Ball State and SMSG programs, the attitude differences for these pupils, although clearly affected by the text difficulty judgments, were not of a similar magnitude. This may reflect in part the stability of the attitudes being measured in general, that is, the tendency for the initial attitudes to be sustained by a number of different factors or conditions. It also may reflect, however, the countering effects of other characteristics of the Ball State and SMSG programs.

For the SMSG program the results suggest that there apparently were other characteristics countering the negative effects resulting from the difficulty pupils had with the materials. The text difficulty differences were of approximately the same magnitude for the Ball State and SMSG programs, however, the SMSG pupils did not show the same tendency to develop less positive attitudes that was observed for the Ball State pupils. Also when E-C comparisons were made adjusting for text difficulty differences (i.e. partialling out the text difficulty effects), there were some indications that with respect to their perception of their own knowledge SMSG pupils developed more positive attitudes than conventional class pupils. No similar tendency was observed for Ball State pupils. It was not evident, however, what qualities of the SMSG program may have contributed to this effect.

Similar differences with respect to text difficulty were not observed for UICSM pupils for whom there was also evidence of the development of somewhat more positive attitudes. However, subsequent analyses suggested that relevant pupil characteristics which were confounded with the instructional treatment differences may have contributed somewhat to these effects.

Among the attitude indices the largest differences occurred on the index of Perceived Utility - a measure of expected usefulness of mathematics for future activities and goals. This measure also showed the lowest pre-post correlations indicating that as a component or dimension of the pupil's attitude or belief concerning mathematics, it was relatively more amenable to change or influence in general than either the pupil's intrinsic interest or his perception of his own ability in mathematics.

The difference in this respect for the pupils' perception of utility may be that this characteristic of mathematics as a subject does not acquire much significance until the ninth grade at which time, among other things, algebra is elected because of its relevance for future educational goals and objectives. Consequently pupils' reactions along this dimension may not be as firmly established on the basis of previous experience as are reactions reflecting intrinsic interest or perceived knowledge.

V. Summary

Questionnaire indices designed to assess pupil attitudes toward and interests in mathematics were administered at beginning and end of the school year to pupils in 38 pairs of ninth grade algebra classes. Each class pair was taught by the same teacher, one class with one of three experimental programs, Ball State, UICSM, or SMSG, the other with a conventional program.

Comparisons were made between E and C class pupils for each of the E programs to determine whether differential attitude changes were obtained on measures of more global attitudes and interest toward mathematics and on measures representing specific mathematics attitude dimensions such as intrinsic interest, perceived utility, and perceived knowledge. Questionnaire measures of other attitude relevant factors such as learning difficulty and judgments of the instructional materials were also obtained as well as mathematics class grades and proficiency test scores.

The results indicated that in general the specific programs of instruction had a relatively small differential effect on pupil attitudes and interest with respect to mathematics as these outcomes were measured in this study.

Among the differences observed was a small but consistent tendency for pupils instructed with the Ball State program to develop less positive attitudes and for those instructed with the UICSM program to develop somewhat more positive attitudes than pupils in the respective comparison classes instructed with conventional programs. Also there was some evidence that pupils instructed with each of the E programs had somewhat more learning difficulty in general than C class pupils. The latter difference was most clearcut for the Ball State pupils.

Comparisons considering instructional factors and conditions relevant to attitude change revealed moderately large differences in "textbook difficulty" for pupils instructed with Ball State and SMSG programs. Among these pupils a larger proportion indicated difficulty understanding their respective textbooks than pupils in the comparison C classes. Additional analysis suggested that this factor was probably contributing to the development of a more negative attitude than would otherwise have been the case for pupils in the Ball State and SMSG classes which appeared to be reflected to a greater extent in the overall attitude differences for the Ball State pupils than for those instructed with SMSG. The latter possibility suggested that with respect to the resultant attitudes there may have been some positive countering effects for the SMSG program. There were no indications of differences in the grades received by the E and C class pupils.

Other analyses indicated that neither variations in initial pupil proficiency as measured by an achievement test, nor the teacher evaluations of, nor their experience with the E programs differentially affected any of the instructional program (E-C) differences.

To determine the nature of these relations in more exact terms and the extent the differences obtained can be replicated for a separate sample, additional analyses will be carried out on this and similar data gathered in a subsequent year.

VI. References

1. Aiken, L. R. "Personality Correlates of Attitude Toward Mathematics," Journal of Educational Research. LVI, 1963. p. 476-480.

2. Atkinson, J. W. An Introduction to Motivation. Princeton, New Jersey: D. VanNosfrand and Co. 1964.

3. Bereiter, C. "Some Persisting Dilemmas in the Measurement of Change," in Hanis, C. W. Problems in the Measurement of Change. Madison: University of Wisconsin Press. 1963. p. 3-20.

4. Brookover, W. B., Thomas, S. and Patterson, Ann. "Self Concept of Ability and School Achievement," Sociology of Education. XXXVII, 1964. p. 271-279.

5. Cronbach, L. J., Rajaratnam, Nagiswari, and Glaser, Goldine C. "Theory of Generalizability: A Liberalization of Reliability Theory," British Journal of Statistical Psychology. XVI, part 2, 1963. p. 137-163.

6. Dutton, W. H. "Attitudes of Junior High School Pupils Toward Arithematic," School Review. LXIV, 1956. p. 18-22.

7. Rorer, L. G. "The Great Response - Style Myth," Psychological Bulletin. LXIII, 1965. p. 129-156.

8. Rosenberg, M. L. "A Structural Theory of Attitude Dynamics," Public Opinion Quarterly. XXIV, 1960. p. 319-340.

9. Ryan, J. J., and Rising, G. R. Participant Teacher Judgments of Experimental Programs in Secondary Mathematics. Unpublished technical report (66-3), Minnesota National Laboratory, Minnesota State Department of Education. 1966.

10. Sequential Tests of Educational Progress, Cooperative Test Division, Educational Testing Service, Princeton, New Jersey. 1957.

11. Winer, B. J. Statistical Principles in Experimental Design. New York: McGraw-Hill. 1962.

12. Castellan, N. J. Jr. "On Partitioning of Contingency Tables," Psychological Bulletin. LXII, 1965. p. 330-338